PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

JUN 14 1933

MAY 1933

Journal of Educational Research

Registered June 4 1929

CONTENTS

		PAGE
Reducing the Variability of Essay Examination Marks Through Eliminating Variations in Standards of Grading	*Verner Martin Sims*	637
The Relation of Personality to Academic Aptitude and Achievement	*Ross Stagner*	648
What Do the Meier-Seashore and the McAdory Art Tests Measure?	*Herbert A. Carroll*	661
A Comparison of Scores on Two College Freshman Intelligence Tests	*Lilian Hunsicker*	666
Do the Stanford-Binet IQ's of Superior Boys and Girls Tend to Decrease or Increase with Age?	*Psyche Cattell*	668
Two Experiments in Learning Educational Statistics	*Dennis H. Cooke*	674
The Gothic Lettering Scale for Grading Slanting Freehand Lettering	*Daryl Branch Miller*	679
Editorial—Mental Hygiene in Education	*Harry J. Baker*	689
New Attacks on Organizing Instruction	*G. A. B.*	691
Research Abstracts and Bibliographies		694
Research News and Communications		701

THE JOURNAL OF EDUCATIONAL RESEARCH is indexed in the Education Index

Copyright 1933, by Public School Publishing Co.

Entered as second-class matter at the Post Office at Bloomington, Illinois, under the act of March 3, 1879

JOURNAL OF EDUCATIONAL RESEARCH

EDITORIAL BOARD

A. S. BARR, *Chairman, Professor of Education, University of Wisconsin, Madison, Wisconsin*

E. J. ASHBAUGH, *Dean, College of Education, Miami University, Oxford, Ohio.* Administration, Supervision, and Teacher Training.

HARRY J. BAKER, *Director, Psychological Clinic, Detroit Public Schools, Detroit, Michigan.* Child Welfare, Guidance, and Development.

LEO J. BRUECKNER, *Professor of Education, University of Minnesota, Minneapolis, Minnesota.* Elementary Education.

CARTER V. GOOD, *Professor of Education, University of Cincinnati, Cincinnati, Ohio.* Teaching, Learning, and Individual Differences.

ERNEST HORN, *Director, University Elementary School and Professor of Education, State University of Iowa, Iowa City, Iowa.* Curriculum Construction.

KAI JENSEN, *Assistant Professor of Education, University of Wisconsin, Madison, Wisconsin.* Preschool Education.

CHARLES W. KNUDSEN, *Professor of Education, George Peabody College, Nashville, Tennessee.* Secondary Education.

CLIFFORD WOODY, *Director, Bureau of Educational Reference and Research and Professor of Education, University of Michigan, Ann Arbor, Michigan.* Measurements, Statistics, and Research.

To save time, manuscripts may be sent directly to members of the editorial board as indicated above.

ASSOCIATE EDITORS

FRED C. AYER, *Professor of Educational Administration, University of Texas, Austin, Texas.*

FOWLER D. BROOKS, *Head of Departments of Education and Psychology, DePauw University, Greencastle, Indiana.*

WILLIAM A. BROWNELL, *Professor of Educational Psychology, Duke University, Durham, North Carolina*

B. R. BUCKINGHAM, *Lecturer in Education, Harvard University, Cambridge, Massachusetts*

W. W. CHARTERS, *Director, Bureau of Educational Research and Professor of Education, Ohio State University, Columbus, Ohio*

VIRGIL E. DICKSON, *Assistant Superintendent of Schools and Director of Bureau of Research and Guidance, Berkeley Public Schools, Berkeley, California*

EDGAR A. DOLL, *Director, Department of Research, Training School, Vineland, New Jersey*

WILLIAM S. GRAY, *Director, Teacher Training, University of Chicago, Chicago, Illinois*

CHARLES E. GREENE, *Assistant Superintendent of Schools, Denver Public Schools, Denver, Colorado*

M. E. HAGGERTY, *Dean, College of Education and Professor of Education, University of Minnesota, Minneapolis, Minnesota*

V. A. C. HENMON, *Chairman, Department of Psychology and Professor of Psychology, University of Wisconsin, Madison, Wisconsin*

KARL J. HOLZINGER, *Professor of Education, University of Chicago, Chicago, Illinois*

EARL HUDELSON, *Dean, College of Education and Professor of Education, West Virginia University, Morgantown, West Virginia*

FREDERICK J. KELLY, *Chief, Section of Colleges and Professional Schools, Office of Education, Department of the Interior, Washington, D. C.*

WILLIAM A. McCALL, *Professor of Education, Teachers College, Columbia University, New York City*

W. J. OSBURN, *Director of Research, State Department of Education, and Professor of School Administration, Ohio State University, Columbus, Ohio*

SHELTON PHELPS, *Dean, Graduate School, George Peabody College, Nashville, Tennessee*

DOUGLAS E. SCATES, *Director, Research and Statistics, Cincinnati, Public Schools, Cincinnati, Ohio*

JESSE B. SEARS, *Professor of Education, Stanford University, California*

FRANCIS T. SPAULDING, *Associate Professor of Education, Harvard University, Cambridge, Massachusetts*

JOHN L. STENQUIST, *Director, Bureau of Educational Research, Department of Education and Lecturer, Johns Hopkins University, Baltimore, Maryland*

PERCIVAL M. SYMONDS, *Associate Professor of Education, Teachers College, Columbia University, New York City*

W. W. THEISEN, *Assistant Superintendent of Schools, Milwaukee, Wisconsin*

M. R. TRABUE, *Executive Secretary of the Committee on Individual Diagnosis and Training, Institute for Employment Stabilization Research, University of Minnesota, Minneapolis, Minnesota*

W. L. UHL, *Dean, School of Education, and Professor of Education, University of Washington, Seattle, Washington*

CARLETON W. WASHBURNE, *Superintendent of Schools, Winnetka Public Schools, Winnetka, Illinois*

GUY M. WHIPPLE, *Secretary, National Society for the Study of Education, Danvers, Massachusetts*

ELIZABETH L. WOODS, *Supervisor, Educational Research and Guidance, Los Angeles City Schools, Los Angeles, California*

G. A. YOAKAM, *Professor of Education and Director, Courses in Elementary Education, University of Pittsburgh, Pittsburgh, Pennsylvania*

THE JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except June, July, and August by the Public School Publishing Company. The subscription price is $3.00; $3.50 in Canada or Mexico; and $3.80 in other foreign countries.

Subscriptions and all communications about reprints or other business matters should be addressed to the PUBLIC SCHOOL PUBLISHING COMPANY, at Bloomington, Illinois.

Address all communications on editorial matters to A. S. Barr, Department of Education University of Wisconsin, Madison, Wisconsin.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXVI MAY, 1933 Number 9

REDUCING THE VARIABILITY OF ESSAY EXAMINATION MARKS THROUGH ELIMINATING VARIATIONS IN STANDARDS OF GRADING

VERNER MARTIN SIMS
University of Alabama

VARIATIONS in the marks assigned by different teachers to the traditional or essay examination have commonly been assumed to result from the subjective nature of the pupil response, but several investigators have called attention to the fact that the variability found may be due not only to the subjectivity of the materials being marked but also to differences in the standards of grading found among school systems and among teachers within a system. The first sort of variation results from "variable" errors, and presumably could be reduced by increasing the objectivity of the marking or by increasing the sampling; but variations of the second sort are indicative of "constant" errors and neither increased objectivity nor increased sampling would necessarily affect them. Before the essay examination can be properly evaluated, certainly before it can be compared with new-type tests, these two kinds of variations must be identified, and, if possible, separated.

The more or less constant nature of the variations in standards of grading suggests that it should be possible to make some correction for them. Starch,[1] in his pioneer work with the essay examination, recognized this fact and by compelling teachers, marking a set of English compositions, to use a scale of 5 points and to conform to the normal probability curve, he was able to reduce the mean deviation from the average of several marks from 4.2 to 2.8, and the average range from the highest to the lowest mark given from 14 to 11.2.

Continued opposition to the arbitrary use of the normal curve, and the possibility of further reduction of the variability by means of a more refined procedure, in addition to the advisability of verifying Starch's

[1] Starch, D. "Can the Variability of Marks Be Reduced?" *School and Society*, II (August, 1915), pp. 242-243.

findings for different types of subject matter and with larger groups, seemed to justify further experimentation.

This paper reports an attempt to eliminate the variability resulting from differences in standards of grading through treating the marks given to essay examination papers as raw scores and by means of a purely objective technique transmuting them to *grades* or school marks.

The data upon which the study was based were: (1) the marks given by twenty-one teachers to a set of twenty-four examination papers in sixth grade arithmetic, reported by Bolton;[2] (2) the marks given by twenty-five teachers to a set of twenty-five examination papers in high school algebra, reported by Shriner;[3] (3) the marks given by twenty-five teachers to a set of twenty-five examination papers in freshman high-school English, also reported by Shriner;[3] and, (4) the marks given by nine readers to a set of twenty examination papers in general psychology, collected by this writer.

The marks given on each of the examinations were reported in percentages, that is, on the basis of 100 points. The readers used by Bolton were teaching sixth grade arithmetic in the schools of Seattle, Washington. Shriner's readers were either teaching, or were advanced students preparing to teach the subject. They were scattered throughout Ohio, Michigan and Indiana. The psychology papers were marked by three instructors in psychology, two 'readers' in the Department of Psychology, and four honor students enrolled in the course at the University of Alabama. The studies of Bolton and Shriner were the only ones found in recent literature where the marks assigned by different readers to a set of papers were actually reported. These marks are not reproduced here, but may be found in the articles referred to above. Since the psychology marks have not been previously reported they are presented in Table I.

If we define objectivity as agreement among competent judges as to the position a paper occupies in a group, the results of these studies indicate a higher degree of objectivity for the essay examination than most of the earlier experimenters have claimed. The average correlation of each reader with the average of all readers for the set of algebra

[2] Bolton, F. E. "Do Teachers' Marks Vary as Much as Supposed?" *Education*, XLVIII (September, 1927), pp. 28-39. Bolton reports the marks of twenty-two teachers, but one record was incomplete and has here been discarded.

[3] Shriner, W. O. "Comparison Factor in the Evaluation of Examination Papers," *Teachers College Journal*, (January, 1930), pp. 65-74.

TABLE I

SHOWING MARKS ASSIGNED BY NINE READERS TO A SET OF TWENTY GENERAL PSYCHOLOGY PAPERS

Student number:	Reader									Range
	A	B	C	D	E	F	G	H	I	
1	84	82	94	74	92	80	87	78	85	20
2	65	62	93	74	72	62	74	69	66	31
3	16	17	49	29	33	32	37	39	21	33
4	68	81	80	64	77	64	76	67	71	17
5	77	68	86	61	85	69	72	73	65	25
6	56	38	67	34	55	17	55	41	31	50
7	47	39	65	43	69	35	45	40	42	30
8	67	60	82	64	83	70	75	68	64	23
9	88	93	96	89	90	93	94	86	88	10
10	76	88	79	77	77	56	83	78	68	32
11	62	58	60	65	76	44	72	49	61	32
12	98	91	95	86	93	95	99	80	89	19
13	96	86	84	77	90	85	81	81	69	27
14	72	69	83	66	85	58	77	60	65	27
15	78	80	87	76	88	81	86	67	62	36
16	51	56	66	62	62	53	71	50	52	21
17	57	52	72	51	78	53	67	58	46	32
18	62	46	63	51	63	45	63	51	50	18
19	57	48	81	58	71	55	75	71	42	39
20	60	61	72	69	75	39	77	58	56	38
Correlation with average mark	.95	.97	.88	.90	.90	.92	.87	.91	.88	
Number failed	6	8	1	6	2	11	3	8	8	

papers was +.946, with a range from +.867 to +.970; for the set of English papers +.917, with a range from +.818 to +.958; for the psychology papers +.91, with a range from +.87 to +.97. Bolton did not report these correlations for his data, and since it is not our purpose here to study the correlation between a reader and the average of all readers we have not bothered to calculate them, but inspection of the several markings shows even higher agreement among the readers of the arithmetic papers.

Such results would seem to indicate that for these readers at least there was decided agreement as to the relative merit of the respective papers. In spite of this seeming objectivity, the variability of the marks given the individual papers by the various readers was extremely great. The average range from the highest to the lowest mark given each of the twenty-four arithmetic papers was twenty-six points and for one paper this range was forty-five points (from 25 to 70). The average range of marks on the algebra papers was twenty-seven, the marks given one paper ranging fifty points (from 39 to 85). The average range for the English papers was twenty-one and for one paper the range was thirty-

nine points (from 39 to 78). The average and maximum range for the psychology papers were 28 and 50 (17 to 67), respectively. When the number of failing marks given by the different teachers is considered, the variability is just as great. For the algebra examination the number of failures assigned by the different teachers ranged from 2 to 16 (8 percent to 64 percent), for the English examination the range of failures was from 1 to 14 (4 percent to 56 percent), and for the psychology examination this range was from 1 to 11 (5 percent to 55 percent). Bolton did not report the passing mark assigned by the teachers who marked the arithmetic papers but there is every reason to believe that the range of failures would have been just as great with this examination.

When one recalls that failures and honors, that our estimates of student ability, both relative and absolute, are commonly based upon the marks assigned by teachers the significance of this variability becomes evident. No matter how high the correlations among different markings may be, so long as such variations in the marks given the same papers by different teachers exist one cannot call the examination a satisfactory measure. To contend that to a certain extent school marks determined from new-type, objective examinations are subject to the same sort of variability is but to beg the question.

Assuming a portion of this variability to be due to variations in the standards of grading, we have attempted to eliminate it through using a uniform procedure for converting the marks of each reader, the scores, to *grades*. Numerous writers have pointed out that the performance of the tested group is the only objective basis possible to use for this purpose. One cannot decide upon what is high performance, or what is low performance, without considering the performance of the group. Elsewhere[4] this writer has proposed an objective method of converting scores to grades which is based upon the central tendency and dispersion of the group scores. The procedure involved locating the passing score at $-1\frac{1}{2}$ standard deviations and then graphing the relationship between scores and grades. The steps in the process are:

1. Locate the lowest passing score. ($-1\frac{1}{2}$ standard deviations).
2. With the lowest passing score as the origin, lay off in equal steps on the vertical axis of a graph the possible scores.
3. With the lowest passing grade used in the particular system as

[4] Sims, V. M. "A Simple Technique for Converting Test Scores to Percentage Grades," *School and Society*, XXXIII (May, 1931), pp. 596-598.

the origin, lay off on the horizontal axis of this graph distances corresponding to the possible grades.

4. Draw a diagonal through the origin and the point corresponding to the maximum possible score and 100 percent.

To find the grade of a particular score: Read across to the diagonal and up or down to the percentage line.

Although designed primarily for use in converting objective test scores to grades, the writer suggested that this technique should be applicable to essay examinations if one but treat the marks given by the readers as *scores* with the maximum possible score one hundred points. If the same lowest passing grade was consistently used, grades thus obtained for the same papers should be free of variations resulting from differences in standards of grading and any variability remaining would presumably be due to the subjectivity of the scoring process.

By means of the procedure outlined above, the marks given by the different readers to the four sets of examination papers mentioned were converted to percentage grades with 70 percent as the lowest passing grade.

To what extent does this conversion reduce the variability? Supposedly with a purely objective test, where each reader's marks were distributed normally, the variations would be completely eliminated. Practically, the results for the four examinations are as shown below.

For each of the examinations the variability of the scores is compared with the variability of the grades by presenting: (1) the range of the scores and the range of the grades assigned to the individual papers by the various readers; (2) the probable error of the scores and of the grades assigned each paper; (3) the variation from reader to reader in the number failed by scores and the number failed by grades. Finally, an attempt is made to interpret and evaluate the variability which remains by determining the amount of actual misplacement (deviation from the average grade) found when the various readers' marks are converted to grades.

In order to show the reduction in the range of the marks assigned the various papers that resulted from changing the scores to grades, frequency distributions of the range of scores and of the range of grades assigned the different papers were prepared for each examination and for the total. These distributions are presented in Table II.

The average range of scores for the total group of papers was 25.5, but this was reduced to 12.3 by converting them to grades. The average reduction for the arithmetic papers was from 26.0 to 11.4, for the algebra

TABLE II

SHOWING DISTRIBUTIONS OF RANGES OF SCORES AND RANGES OF GRADES FOR FOUR SETS OF EXAMINATION PAPERS

Range	Frequency									
	Arithmetic Papers		Algebra Papers		English Papers		Psychology Papers		Total Papers	
	Sc.	Gr.	Sc.	Gr.	Sc.	Gr.	Sc.	Gr.	Sc.	Gr.
50-54.9			1				1		2	
45-49.9	1								1	
40-44.9	3		2						5	
35-39.9	1		4		1		3		9	
30-34.9	2		4		3		6		15	
25-29.9	5		5		4		3		17	
20-24.9	5	1	1	2	6	3	3		15	6
15-19.9	5	5	5	9	6	6	3	2	19	22
10-14.9	1	8	1	9	4	12	1	9	7	38
5- 9.9	1	10	2	4	1	4		8	4	26
0- 4.9				1				1		2
Number	24	24	25	25	25	25	20	20	94	94
Mean	26.0	11.4	27.2	13.5	21.2	13.6	28.3	10.0	25.5	12.3

papers from 27.2 to 13.5, for the English papers from 21.2 to 13.6, and for the psychology papers from 28.3 to 10.0. For the total, for the arithmetic papers, for the algebra papers, and for the psychology papers the average range of scores was greater than the maximum range of grades found for any paper, and only one of the twenty-five English papers had a range of grades as great as the average range of scores.

Although the table does not show it, changing the scores to grades reduced the range for every paper except one, an English paper where the original range was very narrow (five points). It was also found that the amount of reduction (both relative and absolute) was more or less related to the size of the original range. That is, the conversion tended to reduce the large ranges more than the small ones.

These maximum deviations are interesting but they tell but little concerning the trend of the variations. They may well represent extreme cases of variability due to chance factors where the tendency is for most of the marks to cluster closely about an average. To present a more complete picture of the variations found, the probable error (P.E.) of the scores and the probable error of the grades was calculated for each paper. Assuming a normal distribution, this measure, .6745 standard deviations, may be roughly interpreted as follows: Given a particular P. E. for a paper the chances are even that a given mark will lie within 1 P. E. above or below the true mark (the average of an infinite number of marks), the chances are 4½ to 1 that it will lie within 2 P. E.s of the true mark, the chances are 21 to 1 that it will lie within

3 P. E.s of the true mark, the chances are 142 to 1 that it will lie within 4 P. E.s of the true mark, etc.[5]

As a result of converting the scores to grades the probable error for every paper was reduced and for more than one-half of the papers the probable error of the grades was less than one-half as great as that of the scores. Further, the larger probable errors of scores tended to be reduced more (relatively and absolutely) than the small ones.

Frequency distributions of the probable errors of scores, and of grades, for the separate papers of each set and for the total are presented in Table III.

TABLE III

SHOWING DISTRIBUTIONS OF P. E.S OF SCORES AND P. E.S OF GRADES FOR FOUR SETS OF EXAMINATION PAPERS

Probable Error	Frequency									
	Arithmetic Papers		Algebra Papers		English Papers		Psychology Papers		Total Papers	
	Sc.	Gr.	Sc.	Gr.	Sc.	Gr.	Sc.	Gr.	Sc.	Gr.
10.0-10.99							1		1	
9.0- 9.99	1						1		2	
8.0- 8.99										
7.0- 7.99	2		4		1		4		11	
6.0- 6.99	3		3		1		2		9	
5.0- 5.99	2		6		1		8		17	
4.0- 4.99	3		4		7		2		16	
3.0- 3.99	7	3	4	5	7	5	1	3	19	16
2.0- 2.99	4	6	2	13	5	15	1	9	12	43
1.0- 1.99	2	14	2	7	3	4		8	7	33
.0- 0.99		1				1				2
Number	24	24	25	25	25	25	20	20	94	94
Mean	4.3	1.9	4.8	2.3	3.6	2.4	5.9	2.2	4.6	2.2

The reduction following the conversion to grades is obvious from this table. For the total group of papers the average P. E. for the scores was 4.6; for the grades 2.2. For the arithmetic papers the average P. E. of scores was 4.3; and of the grades 1.9; for the algebra papers the two averages were 4.8 and 2.3; for the English papers 3.6 and 2.4; for the psychology papers 5.9 and 2.2. The range of the probable errors of the grades was very narrow. For three of the examinations the average P. E. of scores was greater than the largest P. E. of grades found for any paper, and for the fourth, the English examination, only two of

[5] This measure of variability should not be confused with the *probable error of a score due to the unreliability of the measurement* which is commonly used with objective, particularly standardized, tests. This latter measure is determined by correlating two forms of a test and includes both the error due to subjectivity, if any, and the error due to the limited nature of the sampling.

the 25 papers had a probable error of grades as large as the average probable error of scores.

Since the most serious use made of the mark given on an examination is to determine whether the individual passes or fails, it seemed worth while to make a special study of the number of failures resulting from accepting the teachers' marks at face value compared with the number found when these marks were converted to grades. This comparison was possible for the algebra, English, and psychology examinations only. Bolton, in his published report, did not include information on the number failed by the various readers; consequently, we have not been able to make the comparison for the arithmetic papers.

In Table IV the variation from reader to reader in the number failed when the original marks are accepted is compared with the variations found when these marks are converted to grades. This comparison is made for three of the examinations. In addition, the variation in number failed when the marks are converted to grades is presented for the arithmetic examination.

TABLE IV

SHOWING DISTRIBUTION OF NUMBER FAILED BY SCORES AND BY GRADES FOR READERS OF THREE EXAMINATIONS

No. Failed	Algebra Readers		English Readers		Arithmetic Readers		Psychology Readers	
	Score	Grade	Score	Grade	Score	Grade	Score	Grade
16	1							
15								
14	4		1					
13	1							
12			1					
11	1						1	
10			2					
9	3		3					
8	1		4				3	
7	2		3				1	
6	1		1				1	
5	3	1	5					
4	3	2	2	1				
3	4	5	2	6		2	1	
2	1	9		12		15	1	3
1		8	1	5		4	1	6
0				1				
Number of Readers		25		25		21		9

Not available

The striking fact found here is the decided reduction in the variability from teacher to teacher of number failed, resulting from converting the scores to grades. When the original marks are accepted the number of failures range as follows: for the algebra examination from 2 to 16, for the English examination from 1 to 14, for the psychology examination from 1 to 11. When converted to grades these ranges are 1 to 5, 0 to 4, 1 to 2, respectively. When grades are considered, we find one, two, or three failures given by twenty-two of the twenty-five readers

of the algebra papers, twenty-three of the twenty-five readers of the English papers and by all of the readers of the arithmetic papers. After conversion to grades, six of the nine readers of the psychology papers agreed that only one student should fail, and the three other readers each failed two.

Most investigators are convinced that the essay examination, as ordinarily handled, is unsatisfactory as a measuring instrument. Does the conversion of teachers' marks to grades by the process here outlined make them satisfactory? The answer to this question must of necessity be relative and seems to depend upon the degree of refinement required of the measurement. At best, highly refined measurement (to determine grades at least) is not demanded of the teacher. It is usually assumed that if students can be divided with a fair degree of accuracy into five groups (one failing and four passing) the measurement is satisfactory for all practical purposes of grading. The data previously presented with reference to the probable error of the grades is an indication of their value as measures. The average P. E. found for the ninty-four papers was 2.2, which means that on the average if an infinite number of readers graded a paper the grades would practically all lie within 8.8 points (four P. E.s) of the average of several markings. Stated in another way, the range of 47 points (the lowest grade given any paper was 53, the highest 100) found in the grades used here would mean that the average paper would be rather consistently placed in a given one-fifth of the total range, and would therefore be consistently placed in a five-letter grading system.

These statements are all more or less theoretical, assuming a normal distribution of marks and based upon the nature of this distribution. The small number of readers used for the different examinations, particularly the psychology examination, means that they can be taken as nothing more than rough approximations. In order to show the actual misplacement found for these four examinations we have made one further tabulation. Considering each grade on each paper separately the study is concerned with 1934 grades, given by ninety-four different readers. We have attempted to show the amount of misplacement that would result from accepting the respective readers' *grades* by determining the number of grades that deviate given distances from the average grade given a paper. Cumulative frequency distributions, expressed in percentages, of the deviations from the average grade for the papers of each examination, and for the total, are presented in Table V.

TABLE V

SHOWING CUMULATED FREQUENCY DISTRIBUTIONS OF DEVIATIONS FROM AVERAGE GRADE FOR FOUR SETS OF EXAMINATION PAPERS

Points Deviation	Cumulated Percentage Frequency				
	Arithmetic	Algebra	English	Psychology	Total
0	27.0	24.1	20.5	22.9	23.4
1	55.0	49.5	40.6	44.5	47.6
2	72.1	65.3	59.6	66.7	65.4
3	84.8	75.9	72.6	76.8	77.3
4	89.4	81.8	83.1	83.4	84.4
5	91.8	88.7	89.2	91.7	90.0
6	96.6	92.9	92.6	97.3	94.2
7	98.2	95.8	96.1	100.0	96.9
8	99.2	98.2	97.5		98.4
9	99.4	99.3	98.8		99.2
10	100.0	99.9	99.8		99.8
11		100.0	99.9		99.9
12			100.0		100.0
Number Grades	504	625	625	180	1934

Inspection of the table shows that 65.4 percent of the grades deviate two points or less from the average grade; 90 percent of them deviate five points or less, and 99.8 percent lie within ten points of the average. The differences among the examinations is not pronounced. The psychology examination shows the least range, with arithmetic, algebra and English following in order. The narrow range of the psychology papers may be due to the homogenous nature of the readers used or it may be merely an indication of unreliability because of the small number of readers. The other three examinations are arranged in the order in which one would expect them to fall; that is, one would expect, due to the nature of the subject matter, that the arithmetic examination would show the least variability and the English examination the most. Assuming that deviations of five points or less are not serious, and the forty to fifty points included in the grading scale would in some measure justify such assumption, 90 percent of the students are placed accurately enough for all practical purposes. Deviations of ten points or more are obviously serious, but such extreme deviations are very rare.

In summary, the conversion of these marks to grades by the technique used here reduced the variability as follows:

1. The range of the marks assigned the average paper was reduced more than one-half, from 25.5 to 12.3.
2. The probable error of the marks assigned the average paper was reduced more than one-half, from 4.6 to 2.2.

3. The range in the number failed was decidedly reduced; for the algebra examination from 14 to 4, for the English examination from 13 to 4, and for the psychology examination from 10 to 1. (This information was not available for the arithmetic examination.)
4. Approximately 90 percent of the grades deviate five points or less from the average of several teachers' grades, and a negligible percentage (less than 1 percent) deviate nine points or more.

It is impossible to say whether the grades on these essay examinations represent satisfactory measurement. Surely no one could contend that they are entirely unsatisfactory. Further, if we can assume that we have actually eliminated the constant errors due to differences in standards of grading, the variability of the grades would be decreased as the number of examinations was increased, and one could presumably (although not necessarily practicably) reduce it to a negligible factor. But, assuming an objective procedure for converting scores to grades, new-type, objective examinations are not subject to this sort of error at all, and economy of testing would recommend their use.

It seems to this writer that the case for or against the essay examination depends upon whether it measures abilities that are not (possibly one should say "cannot be") measured by the objective examination. If it actually measures desirable outcomes of instruction that are not measured by the objective test, the evidence presented here seems to indicate that by converting the marks to grades it is possible to get fairly satisfactory measurement with the essay examination. Certainly it seems safe to conclude that if essay examinations are to be used, and, in spite of agitation among testing experts for "objective" examinations, enormous numbers of them are still used, the teachers' marks should be converted to grades.

THE RELATION OF PERSONALITY TO ACADEMIC APTITUDE AND ACHIEVEMENT

ROSS STAGNER
University of Wisconsin

THE prediction of academic achievement is a problem which is still engaging the attention of educational investigators. Although intelligence tests which consistently correlate .50 or better with grades, are now available, attempts at prediction of individual achievement are still quite inaccurate. Two important reasons for the unsatisfactory degree of relationship between aptitude and achievement may be suggested: (1) the defects of methods of grading; and (2) the energy output of the individual student, which probably varies independent of ability. Industry and application will be determined not only by the student's interest in school work, but also by certain personal and emotional traits which influence his activity level. It is the purpose of the present paper to consider the relation of these traits of personality to academic aptitude and achievement.[1] Data from both college and lower educational levels will be mentioned.

The primary difficulty in this problem is the measurement of the personality factor. Aptitude tests are of high reliability and of presumptive validity. Grades, as criteria of achievement, are given and the investigator must use them as they are. Tests for personality traits, on the other hand, are often of low reliability and questionable validity (37). In the face of these difficulties, some relationship between grades, intelligence, and these measures may appear from a consideration of the numerous studies which have been made.

The appended bibliography is annotated with the chief results of previous investigations. It seems desirable, however, to summarize the data for some of the more commonly used measures of personality traits. These include the following: miscellaneous ratings of personality; Pressey X-O test; measures of introversion-extraversion; measures of ascendance-submission; the Downey Will-Temperament test; various psychoneurotic inventories; and the Moss social intelligence test.

[1] These studies have been carried out at the advice and under the direction of Dr. V. A. C. Henmon. My wife, Margaret Stagner, has aided greatly in their completion.

With reference to the results of rating studies there is not much to be said. By increasing the accuracy of instructions and the number of judges, ratings of personal traits may be made of very high reliability. The validity of the result, however, is open to question. To illustrate this point we need mention only the fact that when pupils are rated by the teacher giving their grades, the ratings correlate highly with achievement; when the ratings are by one unfamiliar with their school work, the correlations are low and as often negative as positive. The coefficients reported are often absurdly high (for instance, Webb, (44), gives correlations of .95 to .99 between character traits and grades). The uncritical use of ratings of personal characteristics, then, is likely to lead to the conclusion that these traits are more closely related to grades than is intelligence, or conversely, that they are not related at all. It is probable that neither of these conclusions is correct.

Under certain circumstances the rating method can be used to advantage. Laird (23) measured the correlation of intelligence and achievement in groups selected as having some characteristic in common, *e.g.*, students rated as industrious were compared with those rated as indolent. The results show that personal traits may have a pronounced influence upon the degree to which a person's achievement is determined by his capacity.

The Pressey X-0 test consists of a large number of words from which the student crosses out words which are unpleasant, words denoting things about which he has worried, etc. The total number of words crossed out gives a relative "affectivity" measure, which is alleged to indicate the proportional number of emotional reactions of the individual as compared with the norm groups. Another score is obtained by having the subject encircle the word in each line which has the strongest emotional tone, and computing the number of times he deviates from the choice of the norm group (deviation score). Correlations with both affectivity score and deviation score are low and as often negative as positive (7, 11, 12). Chambers (7) reports a high correlation with a specially derived "differential score," but since this correlation drops considerably when the score is used on a different group, it does not seem valuable for predictive purposes.

The trait "introversion-extraversion" has been measured by the Marston rating scale, the Laird C 1 and C 2 self-rating scales, the Neymann-Kohlstedt questionnaire, the Freyd-Heidbreder questionnaire and the Conklin interests test. The Marston scale is made up of paired

traits, one introverted, the other extraverted, on which the subject is to be rated by himself or by another person. It is reported with correlations of —.21, .26 and —.08 with achievement (14, 15). When age and intelligence are eliminated by the partial correlation technique, Hendrickson (15) reports a correlation for grade-school boys of .46, with achievement. For girls, on the other hand, the correlation is only —.04.

Laird's inventory C 1 has been reported to correlate .35 with grades (18). For the C 2 revision, coefficients are reported of —.26, .00, .11, .24, .30 and .36 (5, 8, 13, 16). When extreme cases are studied, the introverts are found to score higher on intelligence among college students, while in the lower grades extraverts usually score higher on aptitude (15, 17, 28, 45). The introverted college students are reliably higher in achievement than the extraverted group; figures for younger students disagree. When the test scores are correlated with intelligence, the coefficients approximate zero (5, 8, 13, 16, 18).

The Neymann-Kohlstedt, Freyd-Heidbreder and Conklin scales have not been employed in the study of this problem.

The Downey Will-Temperament test attempts to measure traits of temperament and volition through studies of handwriting under various conditions. Its reliability and validity are low. Studies on its relation to achievement have usually been made on "disparity groups," *i.e.*, students achieving more than their intelligence score would predict, as compared with students achieving less. Certain traits have been alleged to go with achievement beyond prediction, but the different studies disagree as to which these traits are (9, 26, 30, 38, 43).

The Allport A-S test measures aggressiveness or self-confidence in face-to-face social situations. It consists of described situations in which the subject is to indicate his probable response. Correlations of .06 and .04 are reported for grades. When extreme cases are studied, one report indicates the superiority of ascendants, the other of submissives (3, 4). Plotting grades against aptitude and ascendance, Broom (4) finds that in proportion to intelligence, dominant students make better grades than those who are submissive.

Beginning with the Woodworth Personal Data Sheet, many questionnaires have been devised to discover the relative frequency of psychoneurotic traits in college students. Laird's B 1 and B 2 inventories employ graphic self-rating scales. Thurstone and Thurstone's Personality Schedule includes 223 questions, to be answered by "yes," "no," and "?" (doubtful). Payne's inferiority scale (14) included fifty statements which the subject marked as true or false relating to himself.

Correlations with grades are .03 for the Woodworth P.D. sheet, .04 and —.10 for the Thurstone, .04 for the Payne scale, .08 for B 1 and .11 and —.10 for B 2 (11, 12, 14, 18, 22, 23, 25, 38, 40, 45). Correlations with intelligence are similarly very low. On the whole it is indicated that neurotic students achieve slightly higher marks than stable individuals. This may be due to the lack of social interests in the neurotic.

The Moss Social Intelligence test shows surprisingly high correlations with grades. Unfortunately, it correlates even more highly with intelligence, and adds little or nothing to the multiple correlation with grades (correlations of .49, .47, .42, .36, .29) (20, 36).

As a sort of summary of this survey of the literature, the following points may be stated:

(1) Objective measures of personality show no linear relationship to either academic aptitude or academic achievement.

(2) There is evidence that unstable, maladjusted students do less well in proportion to their intelligence than do those well-balanced, and that there is a closer correlation of capacity and achievement in stable groups.

(3) Introverts appear to earn proportionately higher marks than extraverts, in a college population.

(4) Dominance also appears to be a factor influencing grades in two groups of equal intelligence.

EXPERIMENTAL DATA

We have made efforts to study this problem further with the use of various objective tests of personality: the Pressey X-0 test (form A), some studies of which have already been mentioned; the ABC questionnaire, a test of emotional stability developed by the author, which has not been standardized; Allport A-S Reaction Study; Laird C 2 inventory; Neymann-Kohlstedt test for introversion-extraversion; Thurstone neurotic inventory; Moss Social Intelligence test; and the Bernreuter Personality Inventory. The available data on the relationship of these tests to academic aptitude and achievement have been reviewed. Most of our data deal with the Bernreuter test, a recent development, which gives scores for neurosis, self-sufficiency, introversion, and dominance. The others were used in preliminary investigations and we are presenting here only the correlations with grades and intelligence.

The populations used were: for the X-0 and ABC tests, 195 upperclass students in the University of Wisconsin; for the Allport, Laird, Moss, Neymann-Kohlstedt and Thurstone tests, 175 upperclass students

TABLE I

RELATION OF PERSONALITY TEST SCORES TO ACADEMIC APTITUDE AND ACHIEVEMENT
(Beneath r is given the number of cases on which it is based. PE's range from .04 to .08.)

Test	r Grades	r Intelligence
X-O "affectivity"	—.049	—.181
	132	60
ABC Questionnaire	—.084	—.254
	153	66
Allport A-S (decile ranks)	—.111	*
	94	
Laird C 2	.247	.187
	174	66
Neymann-Kohlstedt	—.121	—.107
	172	63
Thurstone	—.102	—.094
	148	59
Moss Social Intelligence	.369	.457
	166	62
Bernreuter Inventory:		
Neurosis (men)	.066	.147
(women)	—.127	—.106
Self-sufficiency (men)	.070	.120
(women)	.137	.268
Introversion (men)	.067	.161
(women)	**	
Dominance (men)	—.063	—.097
(women)	.072	.159

* Too few cases to make this correlation of value.
** The introversion score was discarded because we decided that it lacked validity, (correlates .96 with the neurosis score).

at Gustavus Adolphus College; and for the Bernreuter scale, 250 freshmen men and 150 freshmen women at the University of Wisconsin. General intelligence scores for these students are based on the American Council Psychological Examination.

The correlations of these tests of personality with academic grades and with intelligence percentiles are shown in Table I. An inspection will show that this table merely confirms our previous conclusion that no significant relation of a linear nature exists between the variables correlated.

Our further consideration of the data will confine itself to the group of 250 freshmen men, which is the only homogeneous group of sufficient size to break up for statistical analysis.

The first important problem was that of the effect of selecting extreme personality types on average aptitude and achievement. Table II gives the mean intelligence percentile and mean grade-point average for groups scoring one probable error above the mean on personality score, compared with those scoring one P. E. below the mean on the same scale. (This table omits the introversion scale for a reason already noted.)

Table II shows clearly that a considerable difference in aptitude may be nullified by differences in personality. On neurosis the high group has an advantage of 10 percentile points on aptitude, yet there is a difference on achievement of only 0.01 grade points. Similar situations hold for self-sufficiency and dominance. The differences in grades are, however, always in favor of the group having the higher aptitude.

In an attempt to equalize this, we cut out cases scoring above the ninetieth percentile and below the thirtieth. This left forty or more cases in each group, but made the differences in intelligence four points or less. Under these conditions, the mean grade-points were as follows: high N, 1.41, low N 1.50; high S, 1.30, low S, 1.38; high D, 1.35, low D, 1.35. In other words, equalizing the aptitude scores of the groups has reversed their relative achievement for both neurosis and self-sufficiency scales. The stable students and those low on self-sufficiency make higher

TABLE II

APTITUDE AND ACHIEVEMENT OF STUDENTS MAKING EXTREME PERSONALITY SCORES

		Mean	S.D.	Diff.	S.D.$_d$	C.R.
Intelligence:	high N	72.87	20.70	10.86	4.15	2.61
	low N	62.01	27.40			
Grade-points:	high N	1.53	0.77	0.01	0.13	0.07
	low N	1.52	0.79			
Intelligence:	high S	69.98	23.70	8.55	4.27	2.00
	low S	61.43	23.10			
Grade-points:	high S	1.41	0.89	0.05	0.14	0.34
	low S	1.36	0.75			
Intelligence:	high D	64.71	25.30	5.91	4.29	1.37
	low D	70.62	22.26			
Grade-points:	high D	1.41	0.88	0.07	0.14	0.50
	low D	1.48	0.74			

grades despite a slight disadvantage in aptitude. None of these critical ratios is as large as 1.00. This may indicate that the differences are due to chance, or it may indicate that the influence of personality traits is only one of a number of factors determining grades.

The same trend may be shown more clearly by comparing groups from different personality scales. After equalizing intelligence approximately, the low N, high S, and low D scales varied in aptitude by less than one point. The achievement of these groups in order was 1.50, 1.30 and 1.35. Here we have a rather good illustration of the superiority of the emotionally stable student.

Another illustration involves only students of very high aptitude—above the ninetieth percentile. For these students, divided by personality scores, grades were as follows: high N, 1.72, low N, 2.31; high S, 2.00, low S, 2.04; high D, 2.26, low D, 1.80. Each of these is based on

about 12 cases, hence they can only be considered as confirmatory of the indications given in the larger and more heterogeneous group.

In addition to a study of the direct relation of personality scores to aptitude and achievement, it was necessary to find the influence of personality upon the aptitude-achievement correlation. The results are shown in Table III.

It appears from this table that low neurosis, high self-sufficiency and high dominance are conducive of achievement closely proportional to intelligence, while the contrasted traits are likely to lead to achievement either above or below the prediction from ability. This is what we should expect especially from neurotic students, who are likely either to rush into academic work as an escape from social reality, or rush away from school activities because of conflicts associated therewith.

TABLE III

ACHIEVEMENT-INTELLIGENCE CORRELATION WITHIN EXTREME PERSONALITY GROUPS

	r
High Neurosis	.45
Low Neurosis	.60
High Self-sufficiency	.59
Low Self-sufficiency	.37
High Dominance	.71
Low Dominance	.44
Entire group (250 men)	.51

TABLE IV

PERSONALITY SCORES OF "ACHIEVERS" AND "NON-ACHIEVERS"

		Mean	S.D.	Diff.	S.D.$_d$	C.R.
Neurosis:	A's	813.4	63.00			
	Non-A's	835.4	82.44	22.0	15.18	1.44
Self-sufficiency:						
	A's	898.2	42.60			
	Non-A's	921.8	47.40	23.6	9.37	2.51
Dominance:	A's	925.2	50.74			
	Non-A's	918.4	59.42	6.8	11.47	0.59

It was also thought possible that students achieving more than would be predicted from their aptitude score would show differences in personality from those achieving less (comparison of "achievers" and "non-achievers"). This was done by taking the scatter-diagram for the aptitude-achievement correlation and marking off cases in the upper right-hand and lower left-hand corners away from the "line of best fit." Personality scores were then computed for these groups. The differences found are shown in Table IV.

It appears from this table that achievers are characterized by a low emotionality score and a low self-sufficiency score. The differences are large enough to be considered important, if not completely reliable.

GENERAL CONCLUSIONS

From the previous studies which we have reviewed, and from the data just presented, it becomes increasingly clear that personality influences achievement in an indirect way, by affecting the degree to which use is made of the individual's potentialities. This fact explains the uniformly low linear correlations found. At some points along the distribution personality is an advantage in academic work while different amounts of the same personality variable may be disadvantageous, or may be operative in one direction in one case, the opposite in a similar situation.

It is also clear that personality measures can never be used as entrance qualifications, because the students scoring high on undesirable traits are also high on aptitude, and represent the very stratum we should be developing rather than discouraging.

The data are considered an important argument for improved counseling systems in our colleges. It becomes more and more apparent that students of great promise are being allowed to go to wreck upon the reefs of personality maladjustment when a successful adaptation might be relatively simple with adequate guidance. The author's own experience with personal interviews of these students is convincingly in favor of the belief that a few hours of good advice and sympathetic listening to troubles will change the course of their university lives decidedly for the better.

SUMMARY

1. Linear correlations of intelligence, achievement and personality measures are low and are probably so as a result of the inherent nature of the relationship.

2. Extreme personality trends seem to counterbalance advantages in aptitude, making for equal achievement in opposed groups. High emotionality and high self-sufficiency lead to lower achievement that would be predicted from intelligence scores.

3. Personality factors have marked influence on the correlation of aptitude and achievement.

4. Arguments are presented for the use of these measures in improved counseling systems.

BIBLIOGRAPHY

(In the annotations, (1) means correlation of the personality variable with grades, (2) the correlation with intelligence, and (3) means the grade-intelligence correlation reported.)

1. Adams, H. F., Furniss, L., and DeBow, L. A. "Personality as Revealed by Mental Test Scores and School Grades," *Journal of Applied Psychology*, XII (1928), pp. 261-77.
Ratings on 63 traits by teachers: correlate highly with grades, lower with intelligence; (3) 25.

2. Anderson, M. L. "An Investigation of All Pupils of a Defective School in Regard to Their Stability or Instability," *Training School Bulletin*, XXI (1924-25), pp. 129-38.
Stability not related to grades.

3. Bender, I. E. "Ascendance-Submission in Relation to Certain Other Factors in Personality," *Journal of Abnormal and Social Psychology*, XXIII (1928), pp. 137-43.
A-S (1) .06, (2) .00. In extreme cases, submission goes with higher grades.

4. Broom, M. E. "A Study of a Test of Ascendance-Submission," *Journal of Applied Psychology*, XIV (1930), pp. 405-13.
A-S (1) .04, (2) —.012, (3) .39. Dominants get higher grades.

5. Broom, M. E. "A Critical Study of a Test of Introversion-Extroversion," *Journal of Juvenile Research*, XIII (1929), pp. 104-23.
C 2 (two groups): (1) —.26, .00; (2) —.08, .02; (3) .26, .31.

6. Brown, W. M. "Character Traits as Factors in Intelligence Test Performance," *Archives of Psychology*, X, No. 65 (1923), pp. 1-66.
Caution as measured by items skipped on intelligence test: (1) .21; (2) .42.

7. Chambers, O. R. "Character Trait Tests and Prognosis of College Achievement," *Journal of Abnormal and Social Psychology*, XX (1925), pp. 303-11.
X-0 differential score: (1) .54, (2) .23, (3) .33.

8. Craig, H. *Correlation of Laird C 2 with Thorndike Score and Grade-Point Ratio*, Thesis, Stanford University, 1927 (unpublished).
C 2 (1) .36, (2) .07.

9. Downey, J. E. *The Will-Temperament and Its Testing*. Yonkers-on-Hudson, New York: World Book Company, 1923.
Total W-T score (1) .32, (3) .40. Combined ranks on W-T and Thorndike give .55 with grades.

10. Fish, Jean M. *Basal Metabolism, Character Ratings and Academic Grades*. A.B. thesis, University of Wisconsin, 1929 (unpublished).
Ratings by friends on five traits: correlations with grades low, unreliable.

11. Flemming, E. G. "Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen," *Archives of Psychology*, No. 96 (1928), pp. 1-61.
Woodworth Personal Data Sheet, (1) .03, (2) .01, (3) .45. Laird B 2, (1), .11, (2) .01. X-0 total score, (1) .07, (2) —.07. X-0 dev., (1) —.01, (2) .07.

12. Gilliland, A. R., and Voas, W. H. "Personality Tests and Scholastic Achievement." *Personnel Journal*, IX (1930), pp. 305-08.
Laird B 2, (1) —.10, (2) —.24, (3) .43. X-0 total, (1) —.02, (2) —.19. X-0 dev., (1) —.21, (2) —.20.

13. Guthrie, E. R. "Measuring Introversion and Extroversion." *Journal of Abnormal and Social Psychology*, XXII (1927), pp. 82-88.
C 2, (1) .11, (2) .01, (3) .31.

14. Harris, Daniel. "Relation to College Grades of Some Factors Other than Intelligence." *Archives of Psychology*, No. 131 (1931).
Marston rating scale, (1) —.21, (2) —.10, (3) .31.
Inferiority scale (1) .04, (2) .00.

15. Hendrickson, G., and Huskey, J. F. "Extroversion as a Factor Conditioning Achievement in the Fifth and Sixth Grades of the Elementary School." *Journal of Educational Research*, XXV (1932), pp. 6-13.
Marston ratings, Boys (1) .265, (2) .077; girls, (1) —.080, (2) —.056. With age and intelligence constant, (1) boys .460, girls —.046. Age, grades constant, (2) boys —.342, girls —.019.

16. Herrington, L. P. *Scholastic Achievement in a University Group and its Relationship to Measurements of Both the Intellectual and Non-Intellectual Traits*, M. A. thesis, Stanford University, 1928 (unpublished).
C 2 (1) .30, (2) .08.

17. Herrington, L. P. *Physiological Psychology of Introverts and Extroverts*. Ph.D. thesis, Stanford University, 1930 (unpublished).
Extreme introverts, extroverts selected by C 2 and Conklin: I's higher on Thorndike, reliably higher on grades.

18. Hoitsma, R. K. "Reliability and Relationship of the Colgate Mental Hygiene Test." *Journal of Applied Psychology*, IX (1925), pp. 293-303.
C 1, (1) .35, (2) .02. B 1, (1) .08, (2) .00.

19. Hughes, W. H. "Personality Traits and the College Success of High School Graduates." *California Quarterly of Secondary Education*, I (1925-26), pp. 225-36.
Rating studies. Self-confidence negatively related to grades.

20. Hunt, Thelma. "Measurement of Social Intelligence." *Journal of Applied Psychology*, XII (1928), pp. 317-34.
(1) .49 (college upperclass), .47 (college freshmen), .42 (high school).

21. Jasper, H. H. "Measurement of Depression-Elation and its Relation to a Measure of Extraversion-Introversion." *Journal of Abnormal and Social Psychology*, XXV (1930), pp. 307-18.
Elation correlated with higher grades and intelligence (1) .24, (2) .15.

22. Keys, N., and Whiteside, G. H. "Relation of Nervous-Emotional Stability to Educational Achievement." *Journal of Educational Psychology*, XXI (1930), pp. 429-41.

Woodworth-Cady inventory plus ratings: 30 emotional, 30 stable pupils. Stable higher in grades, intelligence.

23. Laird, D. A. "A Study of Some Factors Causing Disparity Between Intelligence and Scholarship in College Students," *School and Society*, XIX (1924), pp. 290-92.
(3) determined for students rated as follows: average, .48; indolent, .34; diligent, —.05; many dates, .13; no dates, .42; lovelorn, .20; engaged, .00; working for board, .53; psychoneurotic, .32; stable, .61 (last two based on B 2).

24. Lauer, A. R., and Evans, J. E. "Influence of a So-Called Emotional Factor on Academic Success," *Journal of Abnormal and Social Psychology*, XXV (1930), pp. 57-59.
Body electric resistance, (1) —.34, (2) —.42. Heart change with exercise, (1) —.65, (2) —.22. (With a second group, (1) —.30.)

25. Mears, C. E. *Interrelationships Between Academic Success, Intelligence and Social Adjustment*, M.A. thesis, University of Wisconsin, 1931 (unpublished).
Superior students (in both intelligence and grades) favored by both principal's ratings, Thurstone inventory, above "Average" group (in both intelligence and grades).

26. Miner, J. B. "Will-Temperament of College Disparity Students," *Psychological Bulletin*, XXII (1925), p. 136.
Certain traits on Downey test may explain achievement above or below ability.

27. Oates, D. W. "Relation of Temperament and Intelligence to Scholastic Ability." *Forum Education*, VII (1929), pp. 171-85.
Rating on temperament, (1) .58, (2) .18, (3) .31.

28. Oliver, R. A. C. "Traits of Introverts and Extroverts," *Journal of Social Psychology*, I (1930), pp. 494-509.
C 2 (10 percentile, 90 percentile): introverts higher on Thorndike, reliably higher on grades.

29. Paynter, R. H., and Blanchard, P. *Study of Educational Achievement of Problem Children*. New York: Commonwealth Fund, 1929.
Grades not related to problem behavior.

30. Poffenberger, A. T., and Carpenter, Florence. "Character Traits and School Success," *Journal of Experimental Psychology*, VII (1924), pp. 67-74.
Downey test: "success traits," care for detail, self-assurance, low motor inhibition, low freedom from load.

31. Pressey, S. L. "Attempt to Measure the Comparative Importance of General Intelligence and Certain Character Traits in Contributing to Success in School," *Elementary School Journal*, XXI (1921), pp. 220-29.
Ratings on "industry and application," (1) .69, (2) .55.

32. Remmers, H. H. *Diagnostic and Remedial Study of Potentially and Actually Failing Students*, Purdue University Bulletin, XXVIII,

No. 12 (Lafayette, Indiana: Purdue University, 1928), pp. 1-164.
Potential failures not emotionally well-adjusted.

33. Rosen, E. K. A Comparison of Neurotic and Normal Children in Public Schools. Contributions to Education, No. 188, New York: Bureau of Publications, Teachers College, Columbia University, 1925.
Not differentiated as to grades.

34. Sangren, P. V. "Social Rating of Best and Poorest High School Students," *Journal of Educational Psychology*, XIV (1923), pp. 209-14.
Rated "scholarship qualities," (1) .93, (2) .85; "citizenship qualities," (1) .92, (2) .77.

35. Stone, C. L. "Disparity Between Intelligence and Scholarship," *Journal of Educational Psychology*, XIII (1922), pp. 241-44.
Disparity group (grade and intelligence ranks differ by 50 percent), speed of writing and (1) .60, (2) .12; association frequency with (1) .00, (2) .18. Normal group (ranks differ by 2 percent), speed of writing with (1) .09, (2) .03; association frequency with (1) .63, (2) .64.

36. Strang, Ruth. "Relation of Social Intelligence to Certain Other Factors," *School and Society*, XXXII (1930), pp. 268-72.
(1) .29, (2) .44. "Test 2" with (1), .15.

37. Symonds, P. M. *Diagnosing Personality and Conduct*, New York: Century Company, 1932.

38. Thurstone, L. L., and Thurstone, T. G. "A Neurotic Inventory," *Journal of Social Psychology*, I (1930), pp. 3-30.
(2) .04. Neurotics, slightly higher grades.

39. Traxler, A. E. "Will-Temperament of Upper-Grade and High School Pupils," *School Review*, XXXIII (1925), pp. 264-73.
(1) .13, (2) .35, (3) .54. Analyze individual profiles for causes of disparity.

40. Turney, A. H. *Factors Other than Intelligence which Affect Success in High School*, (Minneapolis: Minnesota University Press, 1930), pp. 1-135.
Achievers (grades above intelligence rank), compared with non-achievers (below). A's higher on ratings, but not differentiated by tests of introversion, inferiority attitude.

41. Turney, A. H. "Intelligence, Motivation and Achievement," *Journal of Educational Psychology*, XXII (1931), pp. 426-34.
The two major factors in school achievement are intelligence and motivation.

42. Tyler, H. T. *Bearing of Certain Personality Factors Other than Intelligence on Academic Success*, Contributions to Education, No. 468. New York: Bureau of Publications, Teacher's College, Columbia University, 1931.
New rating scale. All correlations with grades low, unreliable.

43. Uhrbrock, R. S. *Analysis of Downey Will-Temperament Tests*,

(New York: Bureau of Publications, Teachers College, Columbia University, 1928), pp. 1-78.
Speed of writing, (1) —.28; freedom from load, (1) .25.

44. Webb, Edward. "Character and Intelligence." *British Journal of Psychology*, No. 3, 1915.
Ratings: *e.g.*, "tendency to do kindnesses on principle," (1) .98.

45. Young, J. B. "How Emotional Traits Predispose to College Failure." *Journal of Educational Psychology*, XVIII (1927), pp. 631-36.
Tests C 2 and B 2. More failures in the extraverted-stable group.

WHAT DO THE MEIER-SEASHORE AND THE McADORY ART TESTS MEASURE?*

HERBERT A. CARROLL
University of Minnesota

A short time before this investigation was undertaken, the writer had found that the correlation between scores earned in the Meier-Seashore Art Judgment Test and scores earned in the McAdory Art Test was but .27 ± .06, a barely reliable relationship. Evidently the two tests do not measure the same kind of art ability.

It was thought that some light might be thrown upon the question as to precisely what each did measure by comparing results with instructors' judgments. Following up this possibility, the writer asked the seven members of the staff of the art education department of the College of Education, University of Minnesota, to arrange in rank order the 103 students who were majoring in art. These instructors were directed to base their ratings on the comparative creative abilities of the students.

Since each instructor was not thoroughly familiar with the abilities of the one hundred or so art majors, it became necessary to make some classification of the basis on which the opinions were given. The following was finally decided upon: (1) comparatively certain; (2) doubtful; and (3) guess. From the scores given, composite ranks were arrived at for those who received comparatively certain and doubtful ratings, and for those who were represented in all three columns. A table was then drawn up showing the percentile ranks of each student according to his instructors' judgment, his success with the McAdory test, and his success with the Meier-Seashore test. (See Table 1). Instructors' judgments in this table include only those which were classified as comparatively certain. The twenty students whose records are presented were chosen at random from the total of 103, the only limitation being that each must have been well enough known to have received a "comparatively certain" rating by at least four of his instructors. Since the instructors numbered seven, four constituted a majority.

An examination of the comparative percentile ranks reveals immediately the marked disagreement among the three measures. For example,

* The writer is deeply indebted to the Committee on Educational Research, University of Minnesota, for initiating and supporting the series of studies in art education of which this is one. Professor Ruth Raymond, head of the art education department, also rendered valuable assistance.

TABLE I

A Comparison of the Percentile Ranks of College Art Students on Art Judgment Tests with Ratings Given by Art Instructors

Students (Key No.)	Instructors N	Instructors P.R. Rating	McAdory Art Test P.R.	Meier-Seashore Art Test P.R.
1.	4	72	70	30
2.	6	88	86	82
3.	4	17	13	14
4.	5	24	40	40
5.	4	28	27	70
6.	4	64	50	17
7.	4	59	48	50
8.	4	31	88	40
9.	4	56	42	60
10.	6	45	42	77
11.	6	94	5	30
12.	7	95	84	40
13.	7	92	70	97
14.	5	72	74	70
15.	7	96	99	99
16.	5	38	61	87
17.	6	64	67	50
18.	4	4	78	30
19.	4	32	72	50
20.	4	41	88	30

number 11 earned a percentile rank of ninety-four from her instructors, but fell to a percentile rank of five on the McAdory Test. On the other hand, in the case of number 15, there is striking agreement.

The resulting correlations appear in Table II. They are not marked, but the Meier-Seashore consistently correlates higher with instructors' judgments than does the McAdory.

TABLE II

Correlations Between Art Instructors' Ratings and Scores on Art Tests

	N	r	P.E.$_r$
Comparatively Certain Ratings			
with Meier-Seashore Art Test	76	.29	.07
with McAdory Art Test	91	.15	.07
Comparatively Certain Ratings + Doubtful Ratings			
with Meier-Seashore Art Test	77	.34	.07
with McAdory Art Test	93	.16	.07
Total of All Ratings			
with Meier-Seashore Art Test	78	.40	.06
with McAdory Art Test	94	.15	.07

EFFECT OF TRAINING UPON ART ABILITY

The writer approached the problem of the effect of training upon art ability in four different ways: (1) through a comparison of the art ability of a group of junior high-school students without art training

with a group of senior high school students with art training; (2) through a comparison of the art ability of these senior high-school students with that of college students; (3) through a comparison of the art ability of college groups classified on the basis of the amount of art education; and (4) through an examination of the amount of improvement made over a one-year period by art majors and by non-art students.

Table III presents the data resulting from the first method of approach. The difference of twelve points in favor of the senior high-school

TABLE III

A Comparison of the Art Judgment Ability of Senior High-School Students with Art Training and Junior High-School Students without Art Training

	N			C.A.				McAdory Art Scores			
	Boys	Girls	Total	Range	Mean	P.E.m	σ	Range	Mean	P.E.m	σ
Senior High-School Art Students	10	25	35	14-11—20-9	17.14	.15	1.31	122-214	173.57	2.13	18.69
Junior High-School Students	10	25	35	11-6—15-0	13.35	.08	.74	104-194	161.29	2.33	20.44
Difference					3.79				12.28		
P.E.D					.17				3.16		
D/P.E.D					22.29				3.89		

art students is not great when it is considered that they are approximately four years older, and that during these four years they had been studying art rather intensively. The junior high-school students, on the other hand, had had no training in art except the little that is ordinarily given in the grades in large city school systems. Whatever it is that the McAdory test measures does not seem to be reliably improved by several years of training on the high-school level.

When, however, these senior high-school students are compared with art majors in college, a reliable difference in art ability is found to exist, although there is considerable question as to whether or not this difference is the result of art training. It may well be that only the abler art students go on to college, or that increase in chronological age accounts for the increase in art judgment ability.

When, on the college level, students who have had a great deal of art training are compared with those of little art training, the difference is negligible, a finding which indicates the conclusion either that training has little effect upon critical ability in art, or that the tests used do not sample materials ordinarily taught in art classes. (See Table IV). There is, however, a reliable difference between the scores of those who

TABLE IV

DIFFERENCES ON MCADORY ART TEST BETWEEN VARIOUS GROUPS SELECTED UPON THE BASIS OF AMOUNT OF ART TRAINING

Amount of Art Training	Students			McAdory Art Test			
	N	Mean	σ	P.E.$_M$	$D_{M_1M_2}$	P.E.$_D$	$^r/_{P.E._D}$
Much Art Training	113	186.90	26.44	1.68			
Little Art Training	96	181.80	22.61	1.56	5.10	2.29	2.22
No Art Training	33	170.20	18.92	2.22	16.70	2.78	6.00
Little Art Training	96	181.80	22.61	1.56			
No Art Training	33	170.20	18.92	2.22	11.60	2.71	4.28

have had much art training and those who have had none. The data do not reveal whether this is due to the training or to a lack of innate ability on the part of those with no training.

The most satisfactory method of measuring the effect of art training upon the kind of art judgment ability measured by tests is obviously to test at the beginning and at the end of a period of instruction. This the writer did with 152 college students, using the Meier-Seashore as a measuring instrument. The results appear in Table V, and are the most significant of any that appear in this discussion. Both the art majors

TABLE V

AMOUNT OF IMPROVEMENT UPON REPETITION OF MEIER-SEASHORE ART JUDGMENT TEST

Subjects	Meier-Seashore Art Test						
	N	Mean	σ	P.E.$_M$	$D_{M_1M_2}$	P.E.$_D$	$^r/_{P.E._D}$
Art Majors							
Dec., 1929	49	82.55	5.68	.55	2.98	.75	3.97
Dec., 1930	49	85.53	5.28	.51			
Non-Art Students							
Dec., 1929	103	72.92	7.77	.52			
Dec., 1930	103	74.52	7.17	.48	1.60	.70	2.28

and the non-art students show a gain over a period of one year. The gain, however, for the non-art students is too small to be reliable. For the art majors, it is 3 points, a very small amount when it is remembered that the group has been subjected during the interval to intensive instruction in the principles of art. The residual gain of art majors over non-art students is 1.38 ± 1.03.

SUMMARY

The first part of this investigation brought to the surface two definite facts: (1) that neither the Meier-Seashore Art Judgment Test nor the McAdory Art Test correlates to any considerable extent with the judgment of university art instructors relative to the creative ability of art majors in college; and (2) that the Meier-Seashore Test agrees with the

instructors somewhat more than does the McAdory, the correlation for the Meier-Seashore and instructors' rating being .40 ± .06 and for the McAdory and instructors' rating, .15 ± .07.

As a result of the second approach made in this study, it was found that training in art appears to influence an art test score but slightly. Whether this is because art tests measure an innate talent rather than a learned reaction, or because they do not measure what is ordinarily taught in art courses is not known.

The writer realizes, as he concludes this paper, that the final impression upon the reader will be badly blurred. The main cause of this is the fact that in the field of appreciation, whether it be art, literature, or music, the investigator has but little to guide or aid him. Measuring instruments are inadequate; objective data few. Once valid and reliable tests have been constructed, innumerable problems in the psychology of aesthetics can be solved. For example, practically nothing is known objectively of the relationship between creative ability and the ability to appreciate; of the relative effect of heredity and environment, of race, of sex, or of factual information upon aesthetic achievement; of the supposed emotional instability of the artistic genius; of the actual value of the kind of training in art and in the other appreciations that is being given in American schools. The door stands wide open to the research worker who is challenged and fascinated by the difficulties involved in a scientific study of such intangible variables.

A COMPARISON OF SCORES ON TWO COLLEGE FRESHMAN INTELLIGENCE TESTS

LILIAN HUNSICKER
Milwaukee, Wisconsin

THE dearth of understanding and the patent misunderstanding of standardized test scores displayed by educators constitute a serious handicap to progress in the scientific study of education. The condition is confined to no one level of teaching; it seems to be universal. College and university faculties surpass, perhaps, the lower schools in furnishing a full quota of educators who show gross ignorance of the elementary tenets of scientific measurement. The interpretation of achievement test scores on the basis of one hundred representing the perfect mark, and a college student's score on an intelligence test as his IQ are typical of the crude fallacies that are far more prevalent than may seem credible to the cloistered laboratory investigator.

The study herein reported was undertaken in the hope that it would be helpful in scattering the fog that so frequently envelops a comparison of two scores of different numerical value, both made by the same subject on different tests of the same trait.

REPORT OF THE STUDY

The Tests Used:

1. Thurstone Psychological Examination, 1929 Edition; administered in the University of Wisconsin state-wide mental survey of high-school seniors, January, 1930.
2. Thorndike Intelligence Examination for High-School Graduates, forms S-L-O, series of 1925-30; administered for the regular freshman intelligence test in the State Teachers College, Milwaukee, Wisconsin, October 1, 1930.

Subjects Who Furnished the Scores:

Those first semester freshmen in the Milwaukee State Teachers College the fall of 1930 for whom there was available a score on each of the two intelligence tests.

TABLE I

SUMMARY OF THE ANALYSIS OF THE TWO DISTRIBUTIONS OF SCORES

Distribution of Scores	N	Range	Mean	Sigma
On Test 1	220	45 to 262	145	48
On Test 2	220	17 to 95	58	15

$r_{12} = .82$, P.E. $\pm$ .01

TABLE II

EQUATING OF THE SCORES ON THE TWO TESTS

Scores Test 1				Test 2
12	13	14	15	17
	16	17	18	18
	19	20	21	19
	22	23	24	20
	25	26	27	21
28	29	30	31	22
	32	33	34	23
	35	36	37	24
	38	39	40	25
	41	42	43	26
44	45	46	47	27
	48	49	50	28
	51	52	53	29
	54	55	56	30
	57	58	59	31
60	61	62	63	32
	64	65	66	33
	67	68	69	34
	70	71	72	35
	73	74	75	36
76	77	78	79	37
	80	81	82	38
	83	84	85	39
	86	87	88	40
	89	90	91	41
92	93	94	95	42
	96	97	98	43
	99	100	101	44
	102	103	104	45
	105	106	107	46
108	109	110	111	47
	112	113	114	48
	115	116	117	49
	118	119	120	50
	121	122	123	51
124	125	126	127	52
	128	129	130	53
	131	132	133	54
	134	135	136	55
	137	138	139	56

Scores Test 1				Test 2
140	141	142	143	57
	144	145	146	58
	147	148	149	59
	150	151	152	60
	153	154	155	61
156	157	158	159	62
	160	161	162	63
	163	164	165	64
	166	167	168	65
	169	170	171	66
172	173	174	175	67
	176	177	178	68
	179	180	181	69
	182	183	184	70
	185	186	187	71
188	189	190	191	72
	192	193	194	73
	195	196	197	74
	198	199	200	75
	201	202	203	76
204	205	206	207	77
	208	209	210	78
	211	212	213	79
	214	215	216	80
	217	218	219	81
220	221	222	223	82
	224	225	226	83
	227	228	229	84
	230	231	232	85
	233	234	235	86
236	237	238	239	87
	240	241	242	88
	243	244	245	89
	246	247	248	90
	249	250	251	91
252	253	254	255	92
	256	257	258	93
	259	260	261	94
	262	263	264	95

By reference to Table II, it is possible to compare, for this group of students, the score on one test with the equated score for the other test. For example: A score of 25 on Test 1, reading across, corresponds to a score of 21 on Test 2; a score of 79 on Test 2, reading backward, corresponds to a score of 213, 212 or 211 on Test 1.

This table has proven helpful, and other tables compiled from similar data, should be useful in the effort to make meaningful to students and instructors a sound interpretation of standardized test scores that differ in numerical value but are trustworthy measures of one and the same trait.

DO THE STANFORD-BINET IQ'S OF SUPERIOR BOYS AND GIRLS TEND TO DECREASE OR INCREASE WITH AGE?

PSYCHE CATTELL
Harvard University

ONE of the important questions for those who administer intelligence tests and use the results in school, clinic, or institution is the question regarding the relative constancy of the intelligence quotients. Early investigations indicated that, on the average, there is a fair degree of consistency, but that some large changes do occur. Later studies have brought out the fact that there is probably a relationship between intelligence level and IQ constancy. The present paper is concerned with the changes which take place with age in the IQ of children of superior intelligence.

In the third volume of *Genetic Studies of Genius*, Professor Terman and his colleagues report the results of retesting eighty-four children who had an initial Stanford-Binet IQ of 140 or higher. The authors found that during the six years between the two tests these children, as a group, lost in IQ. The loss of the boys was only three points, which may not have been significant, but the loss of the girls was thirteen points. This book came into the hands of the writer shortly after the completion of a study in which the results indicated a tendency for the superior child to increase in IQ as he advances in age. Owing to the contradictory conclusions of the two studies, it appeared worthwhile to make a further analysis of the Harvard Growth Study data, especially of the small group of cases with IQ's of 120 or higher.

Only the records of the children who had been given at least two Stanford-Binet tests with a time interval of five or more years between the first and the last were considered, unless otherwise stated. The cases were classified by sex and then into 10-point IQ groups according to the average of the several "corrected" Stanford-Binet IQ's[1] and the medians of the differences between the first and last test calculated for each group. In order to avoid the complicating factor of practice no test was considered unless more than a year had elapsed between it and the preceding test.

[1] The IQ's of those children who passed a total of five or more tests in the 16 and 18 year levels were "corrected" by the method recommended by Professor Terman on page 42 of *Genetic Studies of Genius*, Vol. I.

The results are given in Table I and Figure I. In the two lowest IQ groups the loss between the first and last examination was about 3.5 IQ points, the loss of the boys and girls being almost identical. In the 90 IQ group the boys gained one point while the girls lost two. In all three groups above the 90 IQ group, both boys and girls gained. The higher the IQ level the greater the gain; the boys, however, at each IQ level made a gain which was almost twice that of the girls. In the IQ group of 120 and above the boys made a median gain of twelve points and the girls of seven. The probable error is so large for this group, owing to the small number of cases, that the median would be of little significance if it were not for the general trend of the curves which is too consistent to be ignored.

One of the nine boys in the highest IQ group lost four points; the others gained from seven to twenty-four points each. Of the eleven girls, two lost, one four points and the other six points; the remaining nine gained from two to thirty-four points each. These results are in marked contrast to those of Professor Terman and his colleagues who write:

> "The tabulation from which Table 7 was derived shows that . . . 8 of the total 38 boys lost as much as 15 points each in IQ: . . . 7 of the total of 38 boys gained as much as 15 points each: . . . 16 of the entire 34 girls lost as much as 15 points each; . . . only one of the entire 34 gained as much as 15 points each. There were several cases of extreme change, two . . . boys and four . . . girls each dropping 25 points or more."[2]

In the material from the Harvard Growth Study, four of the nine boys and three of the eleven girls *gained* as much as fifteen points in IQ. None of either sex lost as much as fifteen points, the greatest loss being six points. Two boys and one girl made extreme *gains* of twenty-five or more IQ points.

TABLE I

MEDIAN CHANGE IN IQ AFTER FIVE OR MORE YEARS

IQ Level	Boys No.	Boys Med.	Boys P.E.	Girls No.	Girls Med.	Girls P.E.
Below 80	24	—2.5	1.6	17	—2.5	2.2
80– 89	59	—4.7	1.0	53	—4.3	1.1
90– 99	99	+0.9	0.6	84	—2.0	0.8
100–109	69	+4.3	1.2	70	+2.0	0.9
110–119	28	+6.2	1.8	33	+2.7	1.4
120 and above	9	+12.0	3.6	11	+7.0	3.7
Total	288	+0.97	0.50	268	—0.66	0.50

[2] *Genetic Studies of Genius*, Pages 26-27, Vol. III.

Six of the boys and six of the girls with average IQ's of 120 or higher had been given a third Stanford-Binet examination at least two years after the first test and two years before the last. Unfortunately, the number of cases is too small to bear much weight. It may be of interest, however, to note that in both the boys and girls the average of the middle IQ fell between the first and last IQ. For the girls the averages from first to last were 119, 123 and 132, respectively, and for the boys 119, 133, and 136, respectively.

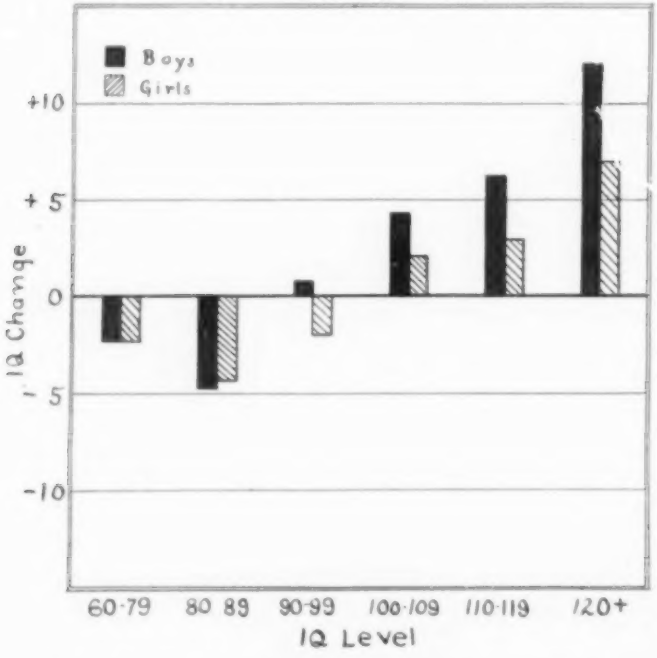


FIGURE I. MEDIAN IQ CHANGES AFTER 5 OR MORE YEARS

There were differences in the selection and treatment of data which probably account, at least in part, for the almost diametrically opposite findings of the two studies.

1. *Examiners:*—Several times as many examiners took part in the Harvard study as in the Stanford study. The small corps of examiners in the Stanford study, probably made for more uniform testing conditions than could be maintained in the Harvard study, but the larger

number of examiners has, at least, one advantage. It has served to eliminate the possibility of the results being modified through a constant error caused by the personal bias of one or two examiners. It is, however, almost certain that such a personal bias, had it been present in the Stanford study, would have been detected. Every test was scored twice from verbatim responses.

2. *Time Interval:*—The time interval between the two tests in the Stanford study was approximately six years in all cases; in the present study the average interval was also six years, but the range was from 60 to 91 months.

3. *Practice:*—There is no indication in the Stanford report that the children were given more than two tests; in the Harvard study nearly half of the "very superior" group had been given six or more examinations. The possibility may occur to the reader that the pupils of the present study may have profited from frequent applications of the tests and thus raised their later IQ's. A recent study gave evidence that the effects of practice on the Stanford-Binet examination does not carry over six months.[3] No case was included in this study in which there had not been a year between the last test and any previous test. Additional evidence on this point was obtained from the records of forty-three children of similar age and IQ level who had repeated the Stanford-Binet examination after a period of from seven to eighteen months. If the increase in IQ reported above had been due to the effects of practice a still larger increase should be present in this group where the tests were given nearer together, but it was found to be only one point.

4. The Stanford cases were selected on the basis of the IQ of 140 or higher on the initial test. The cases in the present study were selected from the Harvard Growth Study material on the basis of a mean "corrected" IQ of 120 or higher from two or more tests. Only one of the Harvard cases obtained an average IQ as high as 140. If there had been a higher group it is, of course, possible that a drop would have been found, but the trend of the curves indicate otherwise.

5. *Selection of Cases:*—A certain amount of lowering of the IQ (the authors indicate three or four points) would be expected from statistical regression in a group of superior children who were selected from their initial IQ as was the Stanford group. This factor, it is believed, was avoided in the Harvard study by giving equal weight to the first and

[3] Cattell, P. "Constant Changes in the Stanford-Binet IQ," *Journal of Educational Psychology*, October, 1931.

last test in the IQ classification. A further study of the Growth study data indicates that the method of selecting cases in the Stanford study may have caused a greater drop in the second IQ than was suspected. Among the Growth study pupils who had had at least two Stanford-Binet examinations with an interval of three or more years between the first and last examinations there were eighty-five cases who had an average IQ of 120 or higher while there were only sixty-seven who entered the study with an IQ of 120 or above. The median of the differences between the first and last examination of the eighty-five cases with an average IQ of 120 or higher was +6.4, while the median of the differences between the first and last test of the sixty-seven cases with initial IQ's of 120 or higher was −0.2. If this sample is comparable with the Stanford group it would account for 6.6 of the nine point drop found in the Stanford study, but since the Stanford group was a much more highly selected group and the time interval between the tests longer it appears possible that a spurious drop, greater than 6.6 points, may be present in the Stanford group. According to the formula given by Professor Kelley,[4] statistical regression would not account for more than about one half of the 6.6 points difference. But the fact remains that when a group of superior children was selected from the same material by each of the two methods this difference in gain was found to be present. It appears possible that whatever the factor, or factors, operating as the cause in the Harvard study may also have been operating in the Stanford study.[5]

6. *Age:*—It appears possible that the difference in the ages of the two groups may have been a contributing cause of the different results found in the two studies. The children with IQ's of 120 and above, whose records were taken from the Harvard Growth Study material, averaged about one year older than those included in the Stanford study, hence were closer to their adult mental level. The Stanford group varied in age from 2-0 to 6-0 years at the time of the first testing and from 8-0 to 13-0

[4] Kelley, T. L. *Interpretation of Educational Measurements.* New York: World Book Company, 1927.

[5] A further study of the trend of the IQ of gifted children is planned at the Psycho-Educational Clinic of Harvard University. In collecting data Dr. E. A. Lincoln has recently retested about sixty children who obtained a Stanford-Binet IQ of 120 or higher at the time they entered the first grade or kindergarten five to eight years ago. These children as a group lost 5.1 IQ points. Those that had an initial IQ of 130 or higher lost 4.3 points, but those that had an *average* of 130 or higher on the two tests, *gained* 7.5. The cases are as yet few in number, there being only twelve cases with an initial IQ of 130 or higher and seventeen who averaged as high, but insofar as they go they corroborate the Growth Study Findings.

years at the second; the range in the Harvard data was from 5-11 to 7-8 and 12-2 to 13-10, respectively. A number of the Stanford group were of pre-school age when first tested, while all of the Growth study pupils were in school. Some psychologists believe that the Stanford-Binet test standards are relatively lower at the bottom of the scale than they are higher up. If this were true it might account for some of the loss in IQ of Professor Terman's group.

Irregularities in the upper end of the scale might cause similar results. We have no Stanford-Binet data from pre-school children, but data obtained from children in the grades indicate that the selection of cases on which the test was originally standardized was a remarkably fortunate one. The number of cases is too few to make a reliable study of age, but, if anything, the older children in the Stanford study have lost more than the younger children.

SUMMARY AND CONCLUSIONS

1. Two studies of gifted children, one made at Stanford University and the other at Harvard University, obtained conflicting results. The Stanford findings point toward a decrease, with age, in IQ of the superior child, while the Harvard Growth Study data point toward an increase.

2. The two factors which appear to be the most probable cause of the conflicting results are: (a) the method of selecting the cases, and (b) the difference in the ages of the two groups.

3. The Harvard Growth Study corroborates that of the Stanford in regard to the direction of the sex difference in the trend of the IQ's of superior children. The Stanford data show a greater decrease in the IQ's of the girls than in those of the boys, while the Harvard data show a smaller gain among the girls than among the boys.

TWO EXPERIMENTS IN LEARNING EDUCATIONAL STATISTICS

DENNIS H. COOKE
George Peabody College for Teachers

FIRST EXPERIMENT

Do students obtain most from courses in educational statistics by spending all of their time in working problems, in reading statistical textbooks and literature, or in a combination of both of these procedures? In order to secure information on this question the writer conducted his course in educational statistics last year in an experimental manner. The eighteen students in the class, all of whom were graduate students, were divided into three groups having equal average intelligence as measured on two group intelligence tests. Upon entering the class only three of the students had any knowledge of educational statistics. One of these students was placed in each of the three groups.

Each student in the reading group was to read two hours daily from a carefully prepared reading list, and prepare a report showing the time and the topics covered. Each student in the problems group was to spend two hours each day working statistical problems. This work was handed in each day. Each member in the combination group was to spend one hour daily reading and one hour in working problems. A check on the amount of time spent on the assignments indicated that each group spent an average of two hours and two minutes daily.

The experiment was conducted according to this plan for six weeks during which time three true-false tests were given. At the end of this period two objective-type achievement tests in educational statistics (prepared by the author) were given, namely, R_I (a test of interpretations) and P_I (a test of processes.) For the second six-weeks of the experiment the reading and problems groups were interchanged, the combination group remaining the same. During this period two true-false tests were given, answers to which were to be given both in and out of class (giving two sets of results). This made a total of seven true-false tests during the experiment. At the end of the second six-weeks period tests R_{II} and P_{II} (separate forms of the test on interpretations and the test on processes, respectively) were given.

The authors of the two intelligence tests used to obtain equivalent groups claim a rather substantial reliability for the respective tests. Fifteen of the twenty-one intercorrelations computed for all of the possible combinations of the scores on the seven true-false tests, are sta-

tistically significant, indicating a certain degree of reliability for the seven tests. The twenty-one r's ranged in size from .27 to .77. The r between forms I and II of the achievement test (R), given at the end of each six-weeks period, is .88 ± .04. The r between forms I and II of the achievement test (P), given at the end of each six-weeks period, is .93 ± .02. These r's indicate a fairly substantial reliability.

RELATIVE ACHIEVEMENT OF THE THREE GROUPS

It may be observed from Table I that the combination group slightly exceeded the reading group on each of the seven true-false tests. The combination group scored higher than the problems group in four of the seven tests, and scored equally with the problems group on one test. The problems group exceeded the combination group on two tests. The reading group scored higher than the problems group in three of the seven tests, while the problems group exceeded the reading group in four tests. In the main, neither the reading nor the problems group exceeded the other in the number of true-false statements scored correctly, each group scoring an average of 304 statements. The combination group, however, scored a slight advantage over both the problems and reading groups, answering an average of 318 statements correctly. Certainly this difference of fourteen statements is not statistically reliable (probably due to the small number of cases involved), but there is a slight tendency for the combination group to exceed both the reading and the problems groups.

TABLE I

MEANS OF THE SCORES ON TRUE-FALSE TESTS* AND RI, RII, PI, AND PII

	True-False Tests							Achievement Tests
	Test II	Test III	Test IV (in class)	Test IV (out of class)	Test V (in class)	Test V (out of class)	Test VI	Scores in RI, RII, PI, and PII
Reading group	41	43	46	49	31	33	61	42
Problems group	36	41	48	55	33	37	54	45
Combination group	42	44	49	53	33	35	62	56

* All students did not take test I; it has, therefore, been omitted.

The problems group slightly exceeded the reading group in average score on the achievement tests (RI, RII, PI, and PII). The combination group exceeded the reading group by fourteen points and the problems group by eleven points. Again, there is a slight tendency for the combination group to exceed both the reading and the problems groups. Due to the small number of cases used, however, the reliability of the differences between the mean scores of the several groups is unknown.

CONCLUSION

On the basis of this small experiment, it does not appear that the teacher of beginning educational statistics should use the reading or the problems method alone, when the combination method appears to give even slightly better results.

SECOND EXPERIMENT

As a result of the first experiment it was concluded tentatively that it is probably more profitable to distribute one's time between problem solving and reading statistical literature (combination method). The best allocation of time between these two procedures was not, however, determined. In order to secure information regarding the best distribution of the student's time between these two types of study the writer conducted a second experiment wherein the thirty students in his class, all of whom were graduate students, were divided into three groups of ten students, each group having equal average intelligence as measured on two group intelligence tests. None of the students had had previous training in educational statistics.

Each student in the reading group was to read for seventy-five minutes daily from a carefully prepared reading list and spend forty-five minutes daily in working statistical problems. Each student in the problems group was to spend seventy-five minutes each day working statistical problems and forty-five minutes in reading statistical literature. Every member in the combination group was to spend sixty minutes daily reading and sixty minutes in working problems. The students turned in to the instructor the problems worked each day and made a daily report showing the topics covered in their reading and the amount of time given to each procedure. A check on the amount of time spent on the assignments daily indicated that each group spent an average of the allotted time plus or minus five minutes.

The experiment was conducted according to this plan for six weeks during which time three true-false tests were given. At the end of this period the same achievement tests in educational statistics, employed in the first experiment, namely, RI and PI, were given. For the second six weeks of the experiment the reading and problems groups were interchanged, the combination group remaining the same. A check on the amount of time given to the assignments daily during the second term indicated that each group spent an average of the allotted time plus or minus four minutes. During this period one true-false test was given,

making a total of four such tests (containing 193 statements) during the experiment. At the end of the second six-weeks period tests RII and PII were given. The reliability of all the tests used was established in the first experiment. It will suffice to say here that the coefficients of correlation were slightly larger in this experiment than in the first.

RELATIVE ACHIEVEMENT OF THE THREE GROUPS

It may be observed from Table II that the combination group slightly exceeded the problems group on each of the four true-false tests. The combination group scored a little higher than the reading group in two of the four tests, scored equally with the reading group on one test, and slightly below the reading group on one test. The reading group scored a little higher than the problems group in three of the four tests, while the problems group slightly exceeded the reading group in one test. For the four tests, the reading group scored a total of 163 statements correctly, while the problems group scored only 157 statements correctly—a difference of six statements in favor of the reading group. The combination group scored a slight advantage over both the problems and reading groups, answering a total of 165 statements correctly. These differences are small, but *again* they favor the combination and the reading groups in the order named.

TABLE II

MEANS OF THE SCORES ON TRUE-FALSE TESTS AND RI, RII, PI, AND PII

	True-False Tests				Total Scores on True-False Tests	Achievement Tests				
Group	I	II	III	IV		RI	RII	PI	PII	Mean
Reading group	15	41	49	58	163	45	49	46	37	44
Problems group	16	39	48	54	157	51	46	51	28	44
Combination group	17	40	49	59	165	50	59	52	40	50

There is no difference in the average score of the problems group and the reading group on the achievement tests, each group scoring an average of 44 points out of a maximum 95. The combination group exceeded the reading and problems groups each by six points. Again, there is a tendency for the combination group to exceed both the reading and the problems groups, but the amount of the difference is small.

To offset the possible advantage that a continuity of process might be to the combination group, the scores may be compared for the first period of the experiment. During the first six weeks of the experiment the reading group scored 105 true-false statements correctly, the problems group, 103, and the combination group, 106. On the two achieve-

ment tests the reading group scored an average of 46 points, the problems group, 51, and the combination group, 51. In view of the small differences there seems to be no choice between the problems and the reading groups, but *again* the combination group either equals or exceeds each of the other groups.

CONCLUSION

Although these data are too meager to warrant any final conclusion, a repetition of the experiment under slightly changed conditions does suggest that an equal distribution of the student's time between working statistical problems and reading statistical literature is slightly preferable to a larger proportion of his time being given either to reading or to working problems.

THE GOTHIC LETTERING SCALE FOR GRADING SLANTING FREEHAND LETTERING

DARYL BRANCH MILLER
San Francisco, California

LETTERING is a subject which presents many difficulties to the person who is endeavoring to put into practice the principles of objective grading. It is essentially a subject that has to be graded subjectively, but it is possible to greatly reduce the reliance upon subjective judgment by the use of a standardized scale. The "Gothic Lettering Scale" has been constructed as a measuring device to evaluate lower case inclined single stroke Gothic lettering or what is commonly called "Reinhardt" lettering, thereby making it possible to apply the principles of objective grading to the subject of lettering.

METHOD OF CONSTRUCTING SCALE

In constructing the "Gothic Lettering Scale" a large number of specimens of lettering were collected from several high schools, from one college, and from a few professional draftsmen. After a preliminary study, about 90 specimens of lettering were selected from the work of students and professionals and were used as a basis for detailed study. Eight of these were picked as key samples. The remaining specimens were roughly grouped into several sets according to merit, then each group was compared with one or more of the key samples. This technique was used instead of the technique of arranging the specimens in order from best to worst because the latter method seems to permit of a great lack of discrimination at the middle of the scale when a large number of specimens are rated. By frequent cross ratings, it is hoped that any bias caused by any peculiar grouping of specimens may have been eliminated.

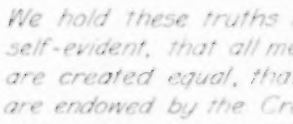
In rating the specimens, the judges were instructed to tell whether those in a given group were better or worse than a given sample. The judges were mechanical drawing teachers, engineering professors, upperclassmen in engineering and in the School of Architecture at the University of California.

After a large number of judgments had been obtained, the number of times that any one specimen was judged better than a nearly related sample was expressed in percentage. This percentage was then converted

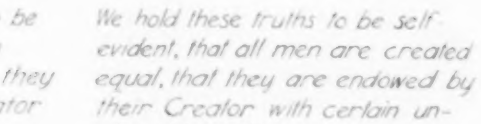
GOTHIC LETTERING SCALE

Numbers 14 and 13 of the scale

14

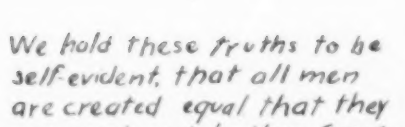

We hold these truths to be
self-evident, that all men
are created equal, that they
are endowed by the Creator
with certain unalienable
rights.

13

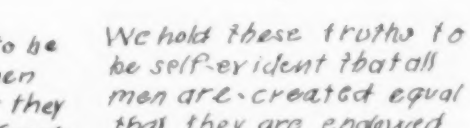

We hold these truths to be self-
evident, that all men are created
equal, that they are endowed by
their Creator with certain un-
alienable rights.

Numbers 7 and 6 of the scale

7


We hold these truths to be
self-evident, that all men
are created equal that they
are endowed by their Creator
with certain unalienable
rights.

6


We hold these truths to
be self-evident that all
men are created equal
that they are endowed
by their Creator with
certain unalienable rights.

into values obtained from the Fullerton Cattell tables for the "Perception of Small Differences." These values were arranged in tables and after numerous arrangements, nine specimens were selected to form the Gothic Lettering Scale.

The Fullerton Cattell values between each pair of adjacent specimens to be selected for the scale are given in Table I. This table gives a critical evaluation of each specimen in the scale. The first column in the table shows the arbitrary scale value assigned to each specimen. The second column gives the serial number of the specimen. In the third column is found the spacing between related pairs of specimens in terms of the Fullerton Cattell values. The mean of all interspacings in column three was computed. Then the difference between this mean and the interspacings of each adjacent pair of specimens was listed in column four. The fifth column gives this difference expressed in percentages with the *mean* taken as a base of 100 percent. This column, then gives the amount of error in the placement of any given specimen and shows how much it varies above or below its ideal position. The last column shows the true position of each specimen in scale values, evaluating from number 10 as a base. From the data in this table, it is possible to critically evaluate the uniformity of the scale.

In the compilation of the data in Table I, it was found that in one instance direct relationships between the specimens could not be found. To overcome this, it was necessary to make indirect calculations by comparing several nearby specimens. The value 2.202 was calculated for the space between 102 and 103.

One further point should be considered as to the accuracy of the scale. It is a matter of chance as to what the next rating on each specimen would have been. Had the next rating been adverse, the position of the specimen might change noticeably in some cases. Upon computing these changes it was found that one adverse rating would not *seriously* change the position of any specimen in an unfavorable direction except specimen 65. In case of specimen 52 at the bottom of the scale, one adverse rating would move it upward close to its correct position.

In order to determine the accuracy with which this scale could be used, an experiment was conducted. A control group of twenty-six specimens was rated eight times with the scale. These ratings were compared with the statistical values of the specimens as had been previously determined when computing the scale.

TABLE I

THE SCALE VALUE OF EACH SPECIMEN IN THE SCALE

Scale Rating	Number of Specimen	Space Between Specimens "Fullerton Cattell" Values	Deviation from Mean	Percent of Deviation	True Scale Value
14	65				14.00
		2.104	—0.288	—12	
13	234				13.12
		2.385	—0.007	0	
12	2151				12.12
		2.492	+0.100	+ 4	
11	218				11.08
		2.575	+0.183	+ 8	
10	167				10.00
		2.211	—0.181	— 8	
9	97				9.08
		2.364	—0.028	— 1	
8	102				8.09
		2.202*	—0.190	— 8	
7	103				7.17
		2.805	+0.413	+17	
6	52				6.00

The Mean space between specimens is 2.392.
* This value was computed.

The results of these eight judgments are shown in Table II. The correlations range from 0.89 to 0.98 with an average of .95. The findings shown in this table indicate that the scale is sufficiently accurate for all practical purposes.

TABLE II

THE RESULTS OF AN EXPERIMENTAL GRADING OF TWENTY-SIX SPECIMENS WITH THE SCALE UNDER CONTROL CONDITIONS

	Mean	Sigma	Coefficient of Correlation between Computed Score and Judged Grades
Computed Score	10.06	2.16	
Mean of 8 Judgments	10.06	2.09	.950
Composite Judgment*	10.04	2.00	.986
Judge No. 1	10.17	1.77	.980
Judge No. 2	9.87	1.95	.962
Judge No. 3	9.85	2.26	.959
Judge No. 4	10.15	2.10	.958
Judge No. 5	9.97	2.03	.977
Judge No. 6	10.10	2.28	.954
Judge No. 7	10.12	2.23	.920
Judge No. 8	10.25	2.07	.890

* The values of the composite judgment were obtained by averaging the eight scores of each specimen as graded by the scale, and then using these values to compute the mean, sigma, and correlation.

CLASSROOM USE OF THE SCALE

Having described the construction of the scale, a few comments should be made concerning the use of the scale. The Gothic Lettering Scale may be used in the classroom as a grading device or as a device for creating motivation in lettering. As a grading device, it is used to insure the teacher in keeping the same basis for grading at all times, not only from month to month, but from year to year. It is next to impossible to keep a uniform standard without some such measuring stick. When an isolated specimen of lettering is graded subjectively, the judge cannot give it its absolute grade, but relates it to what has been judged immediately before. Furthermore, if a single specimen is graded alone at several different times, it is likely to be graded high at one time and low at another time. This is a characteristic of subjective grading.

By using a standardized scale there is always a fixed standard to compare with at every level. Even when using these fixed samples as a basis of comparison there will be irregularity in grading. When using a standardized scale, the irregularities will usually occur only in grading individual specimens. The mean of a class will not vary as a rule far from the true position. When a very accurate grading is desired it is recommended that the specimens be graded three or more times either by different instructors or by the same instructor at different times. The median of these ratings will usually be very close to the correct scale value of a specimen.

To thoroughly evaluate the position of a specimen it should be slowly moved up the scale to a position where it seems to belong and, then again,

it should be moved down the scale to where it seems to belong. This procedure will cause a second evaluation and results in a more accurate discrimination in grading a specimen. When a specimen is similar to a scale sample in style but is really better or worse than the scale sample, this two-fold examination eliminates the bias caused by similarity between styles. The scale when used in the way described above greatly refines the evaluation of the grades in lettering.

As a device for motivation, it has been used by the author to great effect. A copy of the scale was placed on the wall at a convenient place where the students could have access to it at all times. It was found that the students liked to see how well they were doing in comparison with the scale. Often they would go to the greatest pains to try to improve their lettering, so they could attain a higher grade on the scale. It is suggested that teachers place an improvement graph by the side of the scale to further stimulate interest in improvement.

In order to obtain standardized results, it is necessary to use a uniform procedure in giving the test. The first item of importance is to use a standard form for the lettering. The form used in collecting the specimens for the scale required a fixed relation of long letters to short ones. This form can be printed with the stock type of any printing office. The use of a "cut" is unnecessary. The form should be printed with *light* blue or *light* green lines so that the lines will not obscure the lettering. When the results are to be used for statistical study a record blank for the needed information may be printed on the back. The age should be given twice as a check. The form should be printed on a good grade of paper that has been tested to take India ink well.

DIRECTIONS FOR GIVING LETTERING TEST

To obtain satisfactory results in the use of the scale, it is important to induce the students into a coöperative spirit before a test is given. If this is not done, the test will not measure their true ability.

In order to further secure results that may be compared with the standard norms, the following procedure should be used when giving a test.

1. Print the following passage on the blackboard.

   "We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights."

2. Allow the students to practice this passage on scratch paper for five minutes before giving the test. Be sure that the time allowed is exact. Try to avoid giving the impression that this is a timed test, otherwise the students will not feel at ease.
3. Next, pass out slips and have the students fill in the information that is called for on the back of the card. Say to the students, "Use India ink. Keep the ruled side down until you are told to turn the slips over."
4. Then say, "Print the passage that I have written on the board, using lower case letters. Use no capitals except those required by grammar."
5. At the end of five minutes say, "Stop, lay your pens down. I did not expect you to finish."
6. Collect these cards as soon as the ink has dried.

PREPARATION OF STANDARDS OR NORMS

Having constructed the scale, it was necessary to establish standards or norms which would show what grade of work the students actually do in the classroom when measured in terms of the scale. These norms are given in Charts I, II and III. They were constructed from samples of students' lettering which were obtained from eight high schools about San Francisco Bay and from the freshman class in engineering at one of the large western universities. The norms for the freshmen were taken before they had had any drafting in college.

In scoring these, each specimen was graded against the scale by three people who had had experience in drawing and lettering. The median of these three gradings was taken as the score of the specimen. Specimens were graded to half values when necessary. These scores were classified and distributed on a basis of age, the number of terms of drawing that had been taken, and on several other bases as will be described shortly. For each of these distributions the means and standard deviations were computed.

EXPLANATION OF CHART I

Chart I gives the distribution by age for high school students and college students. The ages represented in Chart I are the mid-points. Thus the age of sixteen on the chart represents the age of all students between fifteen years and six months and sixteen years and five months, inclusive. The mean, the standard deviation, the number of cases and the probable error of the mean are found in the enclosed "box" at the bottom

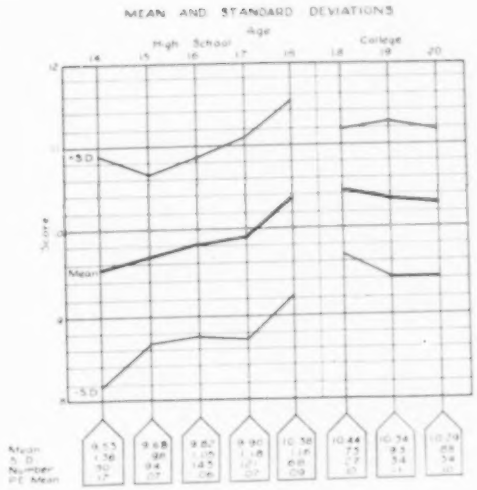


CHART I. AGE NORMS

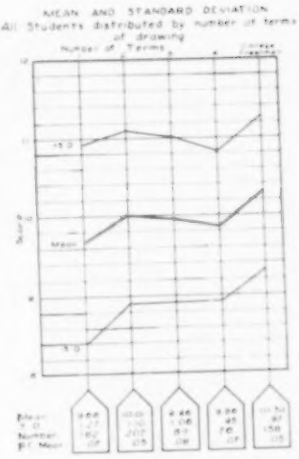


CHART II.
NORMS BASED ON THE NUMBER OF TERMS OF MECHANICAL DRAWING TAKEN

of each graph. The lines in the graph show the position graphically of the mean and standard deviation at each age.

EXPLANATION OF CHART II

Chart II gives the means, standard deviations, etc., for the distribution of high school students classified according to the number of terms that mechanical drawing had been taken. The tests were made at the end of the term. The length of the high school terms was a little over five months. The average age of the students at the end of each term is given below.

	Average Age		
	Years	or	Years and Months
One Term	16.15	or	16–1.8
Two Terms	16.45	or	16–5.4
Three Terms	16.65	or	16–7.8
Four Terms	16.80	or	16–9.6

EXPLANATION OF CHART III

Charts I and II are valuable to one who understands the statistical technique in grading. Chart III has been prepared to aid teachers who are not versed in statistical method. This chart shows a five point grading

system based on findings listed in Charts I and II which may be applied in any classroom.

The graph lines indicate the score at the values of plus one and one-half sigmas, plus one-half sigma, minus one-half sigma, minus one and one-half sigmas, from the mean. The actual percentages of students which fell in each group are indicated on the chart. The ideal statistical percentages in each group are given in the following table.

Excellent Work	7 percent
Good Work	24 percent
Fair Work	38 percent
Unsatisfactory Work	24 percent
Failure	7 percent

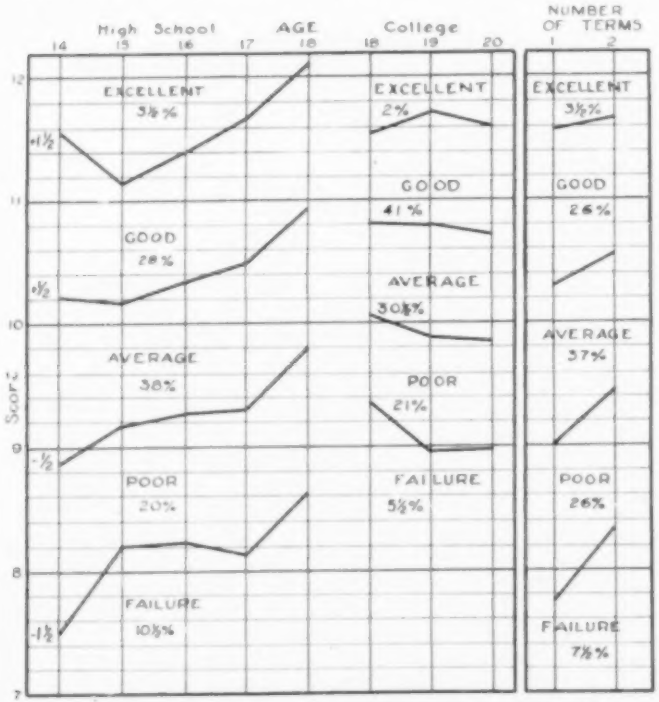


CHART III. A FIVE POINT GRADING SYSTEM: BASED ON VALUES OF 1½, ½, —½, AND —1½ SIGMA

OBSERVATIONS ON CHARTS I, II, III

It will be noticed from these charts that age is an important factor in causing improvement in lettering. When it is remembered that the

average age of students who have taken only one term of drawing is over sixteen years, it is obvious that the younger students are at a decided disadvantage if graded on a basis of class average. Therefore, it seems fair when grading to use the age norms only. This gives the younger students an even chance and also it does not cover up weakness in the work of older students.

FINDINGS OF INTEREST

In collecting the material for the norms, other material was collected which has revealed some interesting facts.

The first fact of interest relates to acceleration and retardation of students. There were 35 accelerated students and 170 retarded students included in Charts I and II.

The range for normal students was placed at three half grades. Thus, a student of the age of 16 years and one month was classed as normal, if he were in the high tenth, the low eleventh, or the high eleventh grade. If this student were in the low tenth, he would be classified as retarded and, if in the low twelfth, he would be classified as accelerated.

It was found in making the distribution of normal students that much greater improvement could be expected of them from age to age than would be indicated in Chart I. Likewise, between the first and second terms the normal students made more progress than the group that contained the retarded students shown in Chart II. After the third and fourth terms, scores of the normal students show little significance as compared to the whole group. This indicates clearly that retarded students do, on an average, inferior work to normal students and their presence materially reduces the class average.

A second interesting observation is that while the quality of lettering increases with age, the quality decreases with the amount of lettering taken after the first year. This would indicate one of three things, either that too little attention is paid to lettering during the third and fourth terms; that lettering becomes burdensome to the student and loses its motivation; or that only the poor students continue in the third and fourth terms. A short investigation conducted along this line indicates that the lack of improvement is due to inadequate motivation.

A last observation was significant as far as the administration of teaching time in the school was concerned. This deals with the influence of giving drawing two periods a day. When all other factors were eliminated, the number of cases was quite small, so that the conclusions

may be doubted, but on the other hand there was a constant tendency in the graphs to act the same way which seemed significant. The student who spent one period a day in drawing seemed to improve almost as much as those who spent two periods a day except during the first term. Therefore, the limited data collected would seem not to justify the time spent in double periods as far as lettering is concerned.

It is regrettable that insufficient data were collected to study the speed element in lettering. It is hoped that at some future time this can be obtained and the norms published.

EDITORIAL

MENTAL HYGIENE IN EDUCATION

If the busy teacher, administrator, or research worker pauses from his labors for a moment and views the whole picture of education he is overwhelmed with the multitudinous activities which beset him on all sides. Everything from new methods of teaching arithmetic, to new systems of ventilation is in process of development, and eventually he is expected to glean the important features, uses, and applications to his daily task of education. Yet among this confusing array we are finding only the beginnings of a movement which is certain to become of great significance—the place of mental hygiene in education.

In self-defense we may believe that this movement is really nothing new, and that we have been practising it regularly for years. While our defense rationalization is true to some extent and without the formal name of mental hygiene, as yet we have not begun to understand its significance, much less to put it into universal practice in the daily routine of the school. The sources of difficulty lie in the misinterpretation and misunderstanding of the actions of children and in our inability to always see problems through the eyes of the child. Through the lack of understanding we pile up more problems because the child is required to accept adult dictums by moral, mental, or physical force.

There are two types of children for whom mental hygiene is required. The first has to do with the normal, average child and his daily program in and out of school; the second deals with the unusual case who unduly tries the patience of the school, upsets its routine, and becomes branded as the blacksheep. The two types differ more in degree than in kind and the fundamental mechanisms of mental hygiene apply to both of them.

In dealing with the average child the classroom teacher must have a sympathetic and understanding point of view toward children if they are to be happy at their work. Two extremes are to be avoided in a balanced mental hygiene program: the one leans in the direction of the strict "school ma'am" type whose word is law, and silence and order are maintained. No one knows what price her pupils pay, possibly in a revolt toward authority in their adult lives which causes them to soon leave any job or position in which there is even a hint of dictatorial autocracy. The other extreme of too great leniency is equally bad, engendering disrespect for any authority or leadership and the development of a false and superficial ego in children. Possibly it is well that

we have as yet practically no long-time follow-up studies to discover such effects. How to mix the right amount of friendliness with the right amount of firmness in dealing with children is an art which needs to be cultivated. We must learn to what extent it can be acquired by teachers and how its potentialities may be measured. The effects should be known of too much coaxing, or the results of ridicule, or of the effects of a continuous flow of empty words of praise and various other practices in daily teaching which are now seldom open to question much less are they the topic of scientific investigation.

The discrepancy between the teacher's point of view on the individual problem child and that of the psychologist and social worker continues to be a stumbling block in the therapeutic program for unusual behavior cases. To the teacher, John, age 12, in the fifth grade, is the sole disturber in the class, he bullies the younger children, he cheats, is noisy and impudent, is tardy and absent more than all of the others combined, and takes no interest in his school work. Nothing, from threats to kindness, appears to have any effect upon him. When viewed from the classroom point of view he is a serious problem, and some drastic and immediate steps need to be taken to cure him or remove him before his class becomes the focal point of chaos. From the standpoint of mental hygiene he is showing what is called "symptomatic" behavior, which means that his actions are the results of certain mental forces and emotions and not merely the outcroppings of native badness or some similar natural tendency to become a hardened criminal. Further analysis will probably show a mild vision defect sufficient to impair active reading and grave enough to lower his vitality to the point of being somewhat nervous. At twelve he is the size of a fourteen year old with the mind of a ten year old which makes for a conflict in his social interests and his actual mental capacity. Added to these physical symptoms he comes from a home in which two older sisters have been model students and he constantly hears so much about it that he hates both his sisters and school as well. His father, who is a successful business man and who attributes his success to hard work rather than to greater ability, does not understand nor sympathize with him, and, as usual, John's mother shields and protects her son. He naturally drifts into the association of questionable companions who are in the same predicament and who therefore can understand him. Eventually he leaves school with no real preparation, and with very gross feelings of injustice. He joins the ranks of those who delay paying their school taxes. His case is by no means the exception; in slightly less acute form there are thousands of such children. They do not belong to special classes; they are within the realm of normal, average children, but their emotional life is badly misunderstood by the school.

There are many implications in John's case. The teacher needs to know the background of such children, and be able to interpret it in terms of the child's feelings and emotions. Either she or some agent

from the school needs to diagnose and interpret this problem to the parents and preferably before the acute emotional problems have developed. By no means has the school done its duty if it contents itself with doing a poor job of teaching him a little arithmetic and reading and then sends him out of the school at the end of the seventh grade with the tangle of emotions and anti-social attitudes which he now possesses.

Problems involving mental hygiene can be solved. They must be interpreted in terms of the teaching situation, the gap must be bridged between what is known about such children clinically and what happens in their daily routine. When added to all the other modern problems mental hygiene applied in education looks like a large order but it is no larger than the challenge which misunderstood children now present to intelligent and progressive educational systems.

HARRY J. BAKER

NEW ATTACKS ON THE PROBLEM OF ORGANIZING INSTRUCTION

THE observer remembers his feeling of shocked surprise at an assertion made by Doctor W. T. Harris about forty years ago. He was discussing the organization of the course of study for the lower elementary grades of the St. Louis Schools and his statement was to this effect: *Chips of knowledge are all we can give to these children.*

If he had said grains of knowledge instead of chips the shock would have been less, for then there would have been the idea of something ready to grow. Even at that time the kindergarten teachers in St. Louis were showing that *chips* did not give the true idea. Teaching in the primary grades has advanced far on the way toward continuity in forty years. Nevertheless we have not always advanced toward the concept of a constantly growing power to learn to use the knowledge and skill nurtured in each pupil.

Continuity is a quality which belongs to the teaching situation for it is essential to the immediate learning situation. Teaching and learning coalesce to become one common situation as pupil and teacher join in that conquest of understanding which ties it up with usable meaning. Of course there is continuity when the usable meaning connects directly with the understanding of how to attack the next difficulty in further learning.

There is, also, another continuity which is built into the school situation as a whole when the organization makes it possible to coördinate the activities of individual initiative with the purposed outcome of class study, and thus connect new attainment with a use of it as an added equipment for personal life.

About twenty years ago, the observer made a series of reports comparing the class work under good teachers each of whom had a different kind or degree of preparation for teaching. He found and reported the

work of (1) the naturally good teacher without formal preparatory training, (2) the able teacher with good normal school training; (3) the teacher of fine artistic power to compose a recitation in which each pupil found an active place and the whole was a unique harmony of all as a working unit which naturally seemed just to grow that way; and finally (4) the genius teacher of wide training and experience who coördinated the activities of the individuals of the class into a complete society of working minds guided by a common purpose.

Research is always required for success in guiding a natural activity. What observable results follow from such an activity? How can the activity itself be measured? When guided, how much of the power is lost in wasted efforts? Such questions are concerned with facts and with their relations to other facts in a way to explain results and to suggest possible changes.

A simple illustration of the kind of facts the research worker in education finds of value comes to the mind of the observer from a frequent experience of his with children less than three years old. As they learn to count the number of things used together, such as the cup, plate, knife, fork, and spoons, used at the dinner table, he has often observed the following situation.

Having been taught the names of the numbers from one to ten or twelve, and learned to repeat them in ordered succession as he handles a like number of blocks; and having but three or four things about his plate, the child will count and recognize that he has numbered these things. But let there be six or seven things and he will repeat the numbers for counting without recognizing that he is to get the number for them all, and will continue to count toward ten until he suddenly recognizes that he is saying numbers for which there are no things before him.

In this observation there are several facts which need to be considered in relation to learning:

1. In learning the names of numbers to ten or twelve, did the child recognize for the first four or five numbers the relation of counting to know if he had four or five blocks? And did he no longer think of counting quantity as he learned the numbers beyond six? Is there any relation to the sense power of the eye or ear of a person to recognize quantity differences up to six without having to count each thing? This power is the basis of the Morse Telegraph code, and may need to be recognized in teaching numbers and perhaps long words to children. It is probable that such natural powers of the eye or ear, and of brain cell recognitions are to be considered with regard to the quickness or slowness of individual responses of children in certain learning situations.

2. In many more complex situations for learning there is a relation of vocabulary or of other preliminary condition, which needs to be frequently checked as a guide to instruction. If part of a vocabulary expresses what is naturally understood, and other words have no roots

of meaning in the child's experiences, then the learning situation and teacher-pupil contacts need to be readjusted for such words.

In the teaching of language, and of the grammatical relations of the parts of speech needed to insure good usage, the observer has found much evidence of inadequate understanding of what use to make of the powers natural to the pupils for expression, and how to meet the special difficulties due to the fact that good usage is, in some instances, not natural to at least a part of the pupils. In this subject there is material for a book on teaching. The same may be said of the teaching of reading. The criticism may be true that at times the pupils are taught how not to read; at least how not to anticipate any satisfaction of a true mental hunger for a community of ideas which may come from reading to understand and appreciate the author through this personal touch with his work.

The artist teacher, referred to above, gained her results by sensing the unity for individual and class organization, and also the necessary order in the presentation of instructional material by which a harmony between thought and action carried the work of the class step by step to the purposed outcome. The project was the buying of the materials and the dressing of a doll to be sold at a bazaar at a price to cover the cost. The costume was designed and the kinds of materials to be used were selected. Notice the two types of arithmetical thinking required. First the amounts of each kind of material must be counted or measured in terms of the selling units of each at the stores. There were some eight different kinds listed as the items on a purchasing list. Some things were measured in inches but must be listed in yards; others were counted by units but were listed in dozens, and so on. Only where the same material was used for different parts of the costume could the measures be added and the sum listed for purchase. Second was the problem of computing the column of prices in dollars and cents, from the unit prices of the different things. Since all items were thus brought under the common unit of price, the column could be added to give the total cost of material.

There were, of course, individual children who had difficulty in thinking the arithmetical principles. The children who could, explained each step and thus aided the teacher to recognize what difficulties needed individual attention. The teacher and the class provided everything and did an artistic piece of work.

The tendency at this time appears to be to put too much of the burden of organization of material and of the means needed for class activity upon the teacher. The research workers are, however, seeking to determine some of the relations of activities for the application of the learning situations required, and to provide the needed means for checking accomplishment and for locating individual difficulties. This work is a necessary basis for teacher training and for the organization of school equipment and supervision.

G. A. B. *Reported by an Observer.*

RESEARCH ABSTRACTS and BIBLIOGRAPHIES

Address all communications relative to research abstracts and bibliographies to A. S. Barr, University of Wisconsin, Madison, Wisconsin.

BULLETINS

BOOGHER, ELBERT W. G. *Secondary Education in Georgia, 1732-1858*, Doctor's Dissertation, University of Pennsylvania (Philadelphia, Pennsylvania: University of Pennsylvania, 1933), vii + 452 pp.

It is the purpose of this investigation to present the facts concerning secondary education in Georgia from the time of the settlement of that colony to the period immediately preceding the Civil War. The bulletin contains a statement of the problem, historical background, early educational beginnings, constitutional and legislative provisions for secondary education, growth of academies, general character of the academies, curriculum and method of teaching, female secondary education, religious influence and denominational schools, and conclusions. Bibliography.

BREGMAN, ELSIE O. *The Performance of Student Nurses on Tests of Intelligence*, The Nursing Education Bulletin, II (New York City: Bureau of Publications, Department of Nursing Education, Teachers College, Columbia University, 1933), 32 pp.

This bulletin presents a report of the performance on standard tests of intelligence of more than 10,000 students of nursing attending accredited schools of nursing in the United States.

CHAMBERLAIN, LEO M., AND MEECE, L. E. *State Performance in Higher Education*, Bulletin of the Bureau of School Service, V, No. 3 (Lexington, Kentucky: College of Education, University of Kentucky, 1933), 37 pp.

The purpose of this study is to bring together certain measures of state performance in higher education in a way that will make possible an easier evaluation of one's state program in the field of higher education in terms of the performances of other states. The data are based upon the last biennial survey of education in the United States.

CHAMBERS, M. M., AND OTHERS. *The First Yearbook of School Law* (Columbus, Ohio: Ohio State University, 1933), 106 pp.

President E. C. Elliot, producing this yearbook, says that "At no point may the school be more profitably studied, as a social and political instrument, than through the judicial determination of its rights and its limitations." The bulletin contains articles relating to the teacher, district boards and officers, county boards and officers, school property, school contracts, district debts, schools taxes, creation and alteration of school districts, etc.

HAGGERTY, M. E., AND OTHERS. *Collegiate Educational Research University of Minnesota:* The Report of the Com-

mittee on Educational Research for the Biennium 1930-1932, Bulletin of the University of Minnesota, XXXVI, No. 1 (Minneapolis, Minnesota: University of Minnesota, 1933), 32 pp.

Some of the projects summarized in this report are: the land grant college study, alumni contacts, utilization of the university plant, relation of the university to the high schools, undergraduate colleges and graduate study, a study of library problems, occupational supply, studies in guidance, freshman curves, probation students, student ratings, mental hygiene of college students, college failures, curriculum studies, instructional improvement, independent study plan, college examinations, comprehensive examinations in mathematics, written examinations in educational psychology, qualifying examinations, studies in physical education, studies in art education, retention of knowledge in psychology, homogeneous grouping, subfreshman English, intensive and extensive reading of essays, student marks in mechanical engineering, integration of course content, and eye movement studies. Bibliography.

HATTENDORF, KATHARINE WOOD, AND OTHERS. *Researches in Parent Education I*, Studies in Child Welfare, University of Iowa Studies, No. 241 (Iowa City, Iowa: University of Iowa, 1932), 288 pp.

The purpose of this investigation centers about the following questions relative to sex education: What should be taught parents? What methods of teaching are effective? How is progress to be measured? And what are the effects of the systematic education of parents? Sample interview forms and incidents offering opportunity for sex instruction are given in the appendix.

HENDERSON, PAULINE, AND OTHERS. *Studies in Measurements and Standardization*, Bulletin on Teaching, XI, No. 3 (Emporia, Kansas: Kansas State Teachers College, 1932), 33 pp.

This bulletin presents brief summaries of eight investigations relating to measurements and standardization.

HORST, AARON PAUL. *A Method for Determining the Absolute Affective Value of a Series of Stimulus Situations*, A Portion of a Doctor's Dissertation, University of Chicago (Chicago, Illinois: The University of Chicago Libraries, 1932), 22 pp.

The purpose of this investigation is to develop a technique whereby a series of stimulus situations can be scaled with reference to their affective value. That is, the degree to which the stimulus situation provokes feelings of pleasantness or unpleasantness.

HUNSBERGER, BYRON K. *Types of Error in Latin Word Knowledge*, Doctor's Dissertation, University of Pennsylvania (Philadelphia, Pennsylvania: University of Pennsylvania, 1932), 246 pp.

The purpose of this investigation is to discover the vocabulary errors of first-year Latin pupils made in a vocabulary test. The bulletin contains a statement of the problem, the tests employed in the investigation, responses associated with Latin, responses associated with English, responses associated with both English and Latin, and the responses not associated with English or Latin. Bibliography.

KEELER, L. W. *Results of the Testing Program in the Branch County Public Schools and in the Michigan State Pub-*

lic School, Bureau of Educational Reference and Research, Bulletin No. 147 (Ann Arbor, Michigan: School of Education, University of Michigan, 1933), 40 pp.

This bulletin presents results obtained from the new Stanford Achievement Test, Primary and Advanced Examination, Form B, administered in the district schools of Branch county and in the Michigan State Public School for dependent children at Coldwater.

LACKEY, EARL E. *Introductory Geography for Teachers'-College Students:* A Method of Selecting and Evaluating Materials for a Professional Subject-Matter Treatment (Lincoln, Nebraska: University of Nebraska, 1933), viii + 124 pp.

This bulletin attempts to set forth the subject matter of an introductory course in geography for teachers'-college students. The bulletin contains a statement of the problem, the technique in arriving at core and marginal materials, an interpretation of the ratings relative to importance, an interpretation of the ratings relative to possibilities of illustrating the involved concepts of geography in the home area, concepts of relationships suggested by the implied questions, a typical unit of professional content, and conclusions. Bibliography.

ROBINSON, FRANCIS P. *The Role of Eye Movements in Reading with an Evaluation of Techniques for Their Improvement*, University of Iowa Studies, No. 39 (Iowa City, Iowa: University of Iowa, 1933), 52 pp.

This investigation is the outcome of certain studies in the Reading Clinic of the University of Iowa. The bulletin contains a statement of the problem, historical background, experimental approach, results of training, the nature of the improvement due to "pacing" eye movements, and summary and conclusions. Bibliography.

SEASHORE, CARL E. *Approaches to the Science of Music and Speech*, University of Iowa Studies, No. 41 (Iowa City, Iowa: State University of Iowa, 1933), 12 pp.

The author sets forth a background of a science of music and speech, pointing out that it is an applied science growing out of contributions from a large number of sciences such as, physics, phonetics, anatomy, physiology, psychology, anthropology, and psychiatry.

SORENSON, HERBERT. *Adult Abilities in Extension Classes:* A Psychological Study (Minneapolis, Minnesota: The University of Minnesota Press, 1933), vi + 100 pp.

The purpose of this investigation was to ascertain some of the psychological and social factors in adult education at the University of Minnesota. The bulletin contains a statement of the purpose and method of the investigation: some characteristics of extension students; relation of college aptitude and reading ability to age, years of schooling, and occupational ranking; comparative abilities of extension students; motives and forces underlying adult education; relative contributions of college ability and years of schooling to achievement; and some interpretations of the findings.

WILLIAMS, HAROLD M., AND OTHERS. *The Measurement of Musical Development*, University of Iowa Studies, VII, No. 1 (Iowa City, Iowa: University of Iowa, 1933), 186 pp.

A large portion of this investigation is given over to techniques which arouse an

interest in young children. This bulletin contains studies in the measurement of musical development, a study of rhythmic performance with special consideration of the factors involved in the formation of a scale for measuring rhythmic ability, and a preliminary study of pitch inflection in the speech of preschool children. Bibliography.

WOODY, CLIFFORD. *Nature and Amount of Arithmetic in Types of Reading Material for the Elementary Schools*, Bureau of Educational Reference and Research, Bulletin No. 145 (Ann Arbor, Michigan: School of Education, University of Michigan, 1932), 79 pp.

This bulletin attempts to throw some light on the contribution of the study of arithmetic to the understanding of other school subjects in the elementary schools. The bulletin contains a statement of the problem, a review of previous investigations and related literature, with a statistical analysis of the results secured. Bibliography.

WOODY, CLIFFORD. *Results of the Testing Program in the Ironwood Public Schools*, Bulletin No. 146, Bureau of Educational Reference and Research (Ann Arbor, Michigan: School of Education, University of Michigan, 1933), 113 pp.

The purpose of this report is to present the results obtained from a testing program administered in the Ironwood Public Schools during the third week of October, 1932. The bulletin contains a statement of the purpose of the investigation, the tests employed in the investigation, and a statistical analysis of the results secured.

EDUCATIONAL ADMINISTRATION AND SUPERVISION

BOLMEIER, E. C. "The Law Governing the Corporal Punishment of Pupils," *Elementary School Journal*, XXXIII (March, 1933), pp. 527-536.

An analysis of legislation relating to the control, management, and correction of pupils, with citations.

CASWELL, HOLLIS L. *Non-Promotion in Elementary Schools*, Field Studies No. 4 (Nashville, Tennessee: Division of Surveys and Field Studies, George Peabody College for Teachers, 1933), x + 100 pp.

This study is a treatment of non-promotion for the guidance of teachers, principals, and superintendents in their study of the problems in their school. The bulletin contains a statement of the purpose and scope of the study, the status of non-promotion in elementary schools, theories of pupil progress, relation of non-promotion to difficulty of instruction, influence of non-promotion on pupils, and solving the problem in individual schools. Bibliography.

CHAMBERS, M. M. "Constitutional Provisions Regarding State Universities," *Educational Law and Administration*, II (April, 1933), pp. 30-38.

The author presents an analysis of constitutional provisions regarding state universities. With specific citations.

ENGLEMAN, JAMES OZRO. "Centralizing Tendencies in Educational Administration in Ohio Since Nineteen Hundred," *Kent State College Quarterly*, XX (February, 1933), vii + 186 pp.

This bulletin contains a statement of the problems, the extent to which educational administration in Ohio was cen-

tralized in 1900, tendencies towards centralized administration in district organization and management, centralized county administration of schools, centralizing administration through centralization and consolidation of schools, steps toward a centralized administration of teacher training, centralized administration in the state department of education, centralized administration of teacher certification, evidences of centralization in the administration of school finance in Ohio since 1900, and conclusions. Bibliography.

PAULU, E. M. "Scientific Supervision in Rural Schools," *Educational Administration and Supervision*, XIX (April, 1933), pp. 241-247.

The author reports upon the use of an informal check list and tests in rural school supervision.

PROFESSIONAL EDUCATION OF TEACHERS

ELIASSEN, R. H., AND ANDERSON, EARL W. "Investigation of Teacher Supply and Demand Reported since November, 1931," *Educational Research Bulletin*, XII, No. 3 (March, 1933), pp. 66-72.

The authors present a digest of the 49 articles on the teacher supply and demand published since November, 1931. Bibliography.

ELIASSEN, R. H. "Pre-training Selection of Teachers," *Educational Administration and Supervision*, XIX (March, 1933), pp. 161-172.

A summary of previous investigations relating to the pre-training of teachers. Bibliography.

TOSTLEBE, M. F. "Analysis of Adequacy of Training Teachers for the One-room Rural School," *Educational Administration and Supervision*, XIX (April, 1933), pp. 260-266.

The author presents an analysis of the reactions of one hundred twenty teachers with one year of training and one or more years of experience, and one hundred eighty cadet teachers to the treatment accorded in the Northern State Teachers College, Aberdeen, South Dakota, to one hundred thirty-five activities common to one-room rural schools.

EXPERIMENTAL STUDIES OF LEARNING AND TEACHING

BUHLER, CHARLOTTE. "The Child and Its Activity with Practical Material," *British Journal of Educational Psychology*, III (February, 1933), pp. 27-41.

The author concludes that the manipulation of materials is of great importance in the intellectual development of children. By the time a child has reached school age he is capable of producing definite results by creative work.

CHANT, S. N. F. "An Objective Experiment on Reasoning," *American Journal of Psychology*, XLV (April, 1933), pp. 282-291.

From an experimental study of reasoning, the author concludes that there are two well-defined methods of reasoning: the first being the interpretative by means of which answers are derived from previously established associations that are remote from the experimental setting itself, and second, the analytical by means of which answers are derived by comparison of items within the experimental setting.

GOODENOUGH, FLORENCE L. "Selected References on Preschool and Parental Education," *Elementary School Jour-*

nal, XXXIII (March, 1933), pp. 541-547.

An annotated bibliography of the most significant publications in the field of preschool and parental education that appeared during the period from April 1, 1932, to December 1, 1932.

LEOPOLD, KATHLEEN B. "The Effect of Creative Work on Aesthetic Appreciation: An Experiment in the Teaching of Poetry," *British Journal of Educational Psychology*, III (February, 1933), pp. 42-64.

While the results are negative, the author concludes from an analysis of individual cases that improvement in poetic judgment may be found even after a short course of training. Certain subjects appear, however, to be quite impervious to aesthetic training.

LORGE, I. "The Efficacy of Intensified Reward and of Intensified Punishment," *Journal of Experimental Psychology*, XVI (April, 1933), pp. 177-207.

From three further experimental studies, the author confirms Thorndike's conclusion that rewards may be depended upon to bring about learning. Improvement is due to the efficacy of rewards in bringing about the repetition of the response so rewarded. A wrong response does more harm by occurring than the punishment can off-set. The positive value of rewards was demonstrated as consistently high and significant.

PESSIN, JOSEPH. "The Comparative Effects of Social and Mechanical Stimulation on Memorizing," *American Journal of Psychology*, XLV (April, 1933), pp. 263-270.

From an experimental study of the comparative effects of mechanical and social conditions on learning non-sense syllables, the author concludes that learning is more efficient under relatively quiet conditions than under conditions of mechanical or social stimulation.

WILLIAMSON, EDMUND G. "The Relation of Learning to Spelling Ability," *Journal of Educational Psychology*, XXIV (April, 1933), pp. 257-265.

From an experimental study of spelling the author found a marked relationship between spelling and intelligence, namely, +.72.

CURRICULUM RESEARCH

GUNDERSON, AGNES G. "Geographical Materials Contained in Readers for the First Three Grades," *Elementary School Journal*, XXXIII (April, 1933), pp. 608-615.

The author presents an analysis of geographical materials contained in ten series of readers used in the first three grades in the elementary school.

KEFAUVER, GRAYSON N., AND MACKENZIE, GORDON N. "The Value of Secondary School Subjects in Preparation for Engineering," *California Quarterly*, VIII (April, 1933), pp. 257-267.

After presenting an analysis of previous investigations the authors make an analysis of entrance requirements of engineering schools and present a survey of the judgment of deans of schools of engineering on the desirability of the different secondary school subjects as preparatory for engineering courses.

SMITH, J. B. "Trends of Thought in Art Education," *School Review*, XLI (April, 1933), pp. 266-277.

The author presents data relative to the significant tendency in the field of art education during the past 40 years.

MEASUREMENTS, STATISTICS, AND THE METHODOLOGY OF SCIENTIFIC RESEARCH

BERNREUTER, ROBERT. "Validity of the Personality Inventory," *Personnel Journal*, XI (April, 1933), pp. 383-386.

The author reports an average reliability of .86 and a validity of .84 and above in the prediction of test scores by other tests of similar traits. Correlations ranging between .56 and .67 were obtained with self-rating.

BROWNELL, WILLIAM A. "On the Accuracy with Which Reliability May Be Measured by Correlating Test Halves," *Journal of Experimental Education*, I (March, 1933), pp. 204-215.

The author concludes from an experimental study of the Spearman-Brown formulae that caution should be used in its application to short objective tests.

CATTELL, PSYCHE. "The Heinis Personal Constant as a Substitute for the IQ," *Journal of Educational Psychology*, XXIV (March, 1933), pp. 221-228.

The author reports data relative to Dr. Heinis' Personal Constant with the conclusion that the Heinis Personal Constant is more constant than the IQ for children below average in intelligence, but for children above average in intelligence it is less constant.

HEILMAN, J. D. "The Translation of Scores into Grades," *Journal of Educational Psychology*, XXIV (April, 1933), pp. 241-256.

The author offers a statistical comparison of three methods of translating test scores into letter grades: a combination of teachers' methods, the 1.6Q-Md method, and the 1.6Q-CMd method. The latter was found to be the most satisfactory. By this method the class median is taken as the mid-value of a grade of *C* and 1.6Q was selected as the range for each grade. The *C* in the formulae is a constant to be added or subtracted to the median for group differences in intelligence.

LINDQUIST, E. F., AND ANDERSON, H. R. "Achievement Tests in the Social Studies," *The Educational Record*, XIV (April, 1933), pp. 198-256.

The authors present certain principles of importance in the construction of achievement tests in the social studies. The principles are illustrated from the 1932 Every-Pupil Test in United States History.

LINDQUIST, E. F., AND COOK, WALTER W. "Experimental Procedures in Test Evaluation," *Journal of Experimental Education*, I (March, 1933), pp. 163-185.

This article presents important evidence concerning the relative effectiveness of certain forms of self-administering spelling tests, demonstrating a new experimental procedure for controlling the time factor in test evaluation.

STAGNER, ROSS. "Improved Norms for Four Personality Tests," *American Journal of Psychology*, XLV (April, 1933), pp. 303-307.

The author gives norms for the Colgate Inventory C2; the Moss Social Intelligence Test, Form 2; Thurstone Neurotic Inventory; and the Pressey X-O Test, Form A.

RESEARCH NEWS and COMMUNICATIONS

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Social Studies Investigation.—The tentative titles of fourteen volumes of the Investigation of the Social Studies in the Schools were listed in the September, 1932, number of this department. To date three of these studies have appeared: *A Charter for the Social Sciences in the Schools*, by Charles A. Beard; *An Introduction to the History of the Social Sciences in Schools*, by Henry Johnson; and *Citizens' Organizations and the Civic Training of Youth*, by Bessie L. Pierce. In addition to the fourteen volumes previously listed, other projected reports will deal with *A Social Process Approach to Curriculum Making in the Social Studies*, by L. C. Marshall; *Experimentation with Social Science Reading Materials*, by C. H. Judd; the history of the social ideas of American educational leaders; freedom of teaching or the growth of tolerance in the teaching of social science; and miscellaneous investigations, many of an exploratory character, made for the commission in charge.

Yearbook on Instructional Materials.—At the meeting of the Executive Committee of the Department of Supervisors and Directors of Instruction in Minneapolis the first of March, it was voted to interpolate a yearbook on "Materials of Instruction" between the yearbook on "The Planning of Supervisory Programs" and the yearbook on "The Supervision of English." This new yearbook, therefore, will be published in February, 1935. There will be a central committee in charge of the yearbook, made up of five persons, with a group of coöperating committees representing interested organizations. Such a yearbook was contemplated by the National Society for the Study of Education, but at the suggestion of the Board of Directors of that organization, the Department of Supervisors and Directors of Instruction has taken up the problem.

General Education Board, 1931-32.—The annual report of the General Education Board for 1931-32 describes its contributions to the fields of natural sciences, humanities, social sciences, science of education, medical sciences, public education, and negro education. The grants toward furthering the science of education were: the sum of $150,000, payable in diminishing amounts over a five-year period to the International Institute of Teachers College, Columbia University, as a supplement to the $100,000 annual contribution over the past ten years; $10,000 to Stevens Institute of Technology to complete and publish a study of aptitude tests for engineering students; $80,000 over a five-year period to Merrill-Palmer School

(Detroit) to support a consultation bureau for college women living in and around Detroit; $9,600 to the Child Guidance Clinic of the University of Rochester to complete the remaining two years of a five-year experiment in the reeducation of problem children; and $10,000 to the Clinic of Child Development of Yale University for the completion of motion pictures depicting early child development and care.

"The chief activity of the Board in the field of public education has continued to be coöperation with the state departments of education of the southern states in their efforts to raise the level of achievement of their respective educational systems. To this end appropriations have been made for studies which would give the needed information on which to build new plans, for the employment of additional personnel in the state departments, for the improvement of teacher-training institutions, and for fellowships for graduate study in education. In addition to the foregoing types of aid restricted to the southern states, the Board gave funds for the completion of a national study of state school finance."

The state departments of education aided were: Alabama, Georgia, Mississippi, Tennessee, and Virginia. Grants to teacher-training institutions were: $150,000 to the University of Florida toward a laboratory high school building; $1,000 for each of two years toward salaries in the practice school at the University of South Carolina; and $7,500 to George Peabody College for Teachers for the department which deals with problems involved in the separate education of white and colored groups in the southern states. Grants for specific investigations were: $5,000 toward a survey of education in Kentucky; $7,500 toward a study of curricular revision in Virginia; $20,000 for a study concerned with coördination of state institutions of higher education in Georgia; and $25,000 to the American Council on Education to complete a study of state systems of school support, begun as a part of the nation-wide study of school finance, the latter of which was discontinued due to curtailment of funds by Congress.

In the field of negro education, with the exception of emergency aid, the Board has sought during the past year mainly to assist in carrying out those projects to which it was already committed. Among the institutions aided were: Atlanta University Library, an endowment of $600,000; Atlanta School of Social Work, $10,000 yearly for three years; Dillard University, $12,500 for each of two years; Fisk University, $75,000 over two years; LeMoyne College, $43,000; Morgan College, $4,500; Paine College, $1,500; and Bacone College, $65,000. Emergency aid to other negro institutions totaled more than $177,000. The Board gave $6,000 toward a survey of Methodist-Episcopal Negro Colleges.

During the year appropriations were $3,908,846 from income and $6,907,300 from principal.

Inglis Lectures.—One of the most outstanding series of critiques of secondary education in print is the Inglis Lectures delivered at and published by the Harvard Graduate School of Education. The authors, titles, and dates of publication are as follows:

Leonard V. Koos, "Trends in American Secondary Education," 1925.

Paul H. Hanus, "Opportunity and Accomplishment in Secondary Education," 1926.

Abraham Flexner, "Do Americans Really Value Education?" 1927.

Charles H. Judd, "The Unique Character of American Secondary Education," 1928.

George S. Counts, "Secondary Education and Industrialism," 1929.

Thomas H. Briggs, "The Great Investment: Secondary Education in a Democracy," 1930.

John Dewey, "The Way Out of Educational Confusion," 1931.

William Setchel Learned, "Realism in American Education," 1932.

H. C. Morrison, "The Evolving Common School," 1933.

Report of the Survey of Secondary Education.—In the letter of transmittal which appears in Monograph No. 22 of the National Survey of Secondary Education, Commissioner Cooper makes a very pertinent statement concerning the organization and scope of the survey:

"Within a period of thirty years the high-school enrollment has increased from a little over 10 percent of the population of high-school age to more than 50 percent of that population. This enrollment is so unusual for a secondary school that it has attracted the attention of Europe, where only 8 to 10 percent attend secondary schools. Many European educators have said that we are educating too many people. I believe, however, that the people of the United States are now getting a new conception of education. They are coming to look upon education as a preparation for citizenship and for daily life rather than for the money return which comes from it. They are looking upon the high school as a place for their boys and girls to profit at a period when they are not yet acceptable to industry.

"In order that we may know where we stand in secondary education, the membership of the North Central Association of Colleges and Secondary Schools four years ago took the lead in urging a study. It seemed to them that it was wise for such a study to be made by the Government of the United States rather than by a private foundation; for if such an agency studied secondary education it might be accused either rightly or wrongly of a bias toward a special interest. When the members of a committee of this association appeared before the Bureau of the Budget in 1928 they received a very courteous hearing. It was impossible, so the chief of the Budget Bureau thought, to obtain all the money which the commission felt desirable; with the money which was obtained, $225,000, to be expended over a three-year period, it was found impossible to do all the things that the committee had in mind. It was possible, however, to study those things which pertained strictly to secondary education, that is, its organization; its curriculum, including some of the more fundamental subjects, and particularly those subjects on which a comparison could be made between the present and earlier periods; its extracurriculum, which is almost entirely new in the past thirty years; the pupil population; and administrative and supervisory problems, personnel, and activities.

"The handling of this survey was entrusted to Dr. Leonard V. Koos of the University of Chicago. With great skill he has, working on a full-time basis during his off-quarters from the University of Chicago, and part-time the other quarters, brought it to a conclusion."

The titles, authors, and prices of the twenty-eight monographs (published by the Office of Education and for sale by the Superintendent of Documents, Wash-

ington, D. C.,) in which the full report of the survey appears are as follows:

1. "Summary," Leonard V. Koos and Staff. 15 cents.
2. "The Horizontal Organization of Secondary Education—A Comparison of Comprehensive and Specialized Schools," Grayson N. Kefauver, Victor H. Noll, and C. Elwood Drake. 20 cents.
3. "Part-Time Secondary Schools," Grayson N. Kefauver, Victor H. Noll, and C. Elwood Drake. 10 cents.
4. "The Secondary-School Population," Grayson N. Kefauver, Victor H. Noll, and C. Elwood Drake. 10 cents.
5. "The Reorganization of Secondary Education," Francis T. Spaulding, O. I. Frederick, and Leonard V. Koos. 40 cents.
6. "The Smaller Secondary Schools," Emery N. Ferriss, W. H. Gaumnitz, and P. Roy Brammell. 15 cents.
7. "Secondary Education for Negroes," Ambrose Caliver. 10 cents.
8. "District Organization and Secondary Education," Fred Engelhardt, William H. Zeigel, Jr., William M. Proctor, and Scovel S. Mayo. 15 cents.
9. "Legal and Regulatory Provisions Affecting Secondary Education," Ward W. Keesecker and Franklin C. Sewell. 10 cents.
10. "Articulation of High School and College," P. Roy Brammell. 10 cents.
11. "Administration and Supervision," Fred Engelhardt, William H. Zeigel, Jr., and Roy O. Billett. 15 cents.
12. "Selection and Appointment of Teachers," W. S. Deffenbaugh and William H. Zeigel, Jr., 10 cents.
13. "Provisions for Individual Differences, Marking, and Promotion," Roy W. Billett. 40 cents.
14. "Programs of Guidance," William C. Reavis. 10 cents.
15. "Research in Secondary Schools," William H. Zeigel, Jr. 10 cents.
16. "Interpreting the Secondary School to the Public," Belmont Farley. 10 cents.
17. "The Secondary-School Library," B. Lamar Johnson. 10 cents.
18. "Procedures in Curriculum Making," Edwin S. Lide. 10 cents.
19. "The Program of Studies," A. K. Loomis, Edwin S. Lide, and B. Lamar Johnson. 15 cents.
20. "Instruction in English," Dora V. Smith. 10 cents.
21. "Instruction in the Social Subjects," William G. Kimmel. 10 cents.
22. "Instruction in Science," Wilbur L. Beauchamp. 10 cents.
23. "Instruction in Mathematics," Edwin S. Lide. 10 cents.
24. "Instruction in Foreign Languages," Helen M. Eddy. 10 cents.
25. "Instruction in Music and Art," Anne E. Pierce and Robert S. Hilpert. 10 cents.
26. Nonathletic Extracurriculum Activities," William C. Reavis and George E. Van Dyke. 15 cents.
27. "Intramural and Interscholastic Athletics," P. Roy Brammell. 10 cents.
28. "Health and Physical Education," P. Roy Brammell. 10 cents.

THE TEACHING OF GEOGRAPHY

The Thirty-Second Yearbook of the National Society for the Study of Education

Prepared by the Society's Committee on the Teaching of Geography

Robert M. Brown, Earl E. Lackey, Edith Parker, Douglas Ridgley, DeForest Stull, Zoe Thralls, W. C. Bagley, and A. E. Parkins (Chairman)

Assisted by Numerous Members of the Society

This great yearbook was made possible by stout and continued assistance and cordial support of the National Council of Geography Teachers. Proposed originally by Miss Flora Nettleman of Toledo, fostered further by Dr. Ernest Horn and Dr. W. C. Bagley, the undertaking began in May, 1930, under the direction of the Society's Yearbook Committee on the Teaching of Geography composed of the following: Professor A. E. Parkins (chairman), Professors R. M. Brown, Earl E. Lackey, Edith Parker, Douglas Ridgley, DeForest Stull, Zoe Thralls, and W. C. Bagley.

Dr. Guy M. Whipple, Secretary of the National Society for the Study of Education and Editor of the Society's Yearbooks, says in his Editor's Preface, "Seldom, if ever, has the Society produced a yearbook that represents so nearly a *consensus gentium*—the *gentes* here being that large group of individuals professionally concerned with the place of geography in the program of education."

The titles of the six sections are:

Section I. The Development of Geography and Its General Contribution to Life
Section II. Developing the Science of Teaching Geography
Section III. Geography in the Curriculum
Section IV. The Curriculum in Geography
Section V. The Teaching of Geography
Section VI. Investigations in the Teaching of Geography

Bound in one volume of 615 pages

Cloth covers at $4.50 Paper covers at $3.00

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS

ACTIVE ELEMENTS THAT MUST BE COORDINATED TO ATTAIN SUCCESS IN TEACHING

(1) the teacher
(2) the textbook
(3) the pupil's learning power
(4) the testing that measures the development of a working use of the pupil's abilities

(1) The Teacher

For summer teacher-training classes and for meetings of teachers-in-service during the school year we recommend the fifteen numbers of the Educational Problem Series in the fields of elementary school instruction, secondary education, school management, educational psychology, rural education, and logic.

Of the Yearbooks of the National Society for the Study of Education we recommend particularly:

The Teaching of Geography (32d Yearbook)
The Teaching of Science (Pt. I, 31st Yearbook)
The Teaching of Arithmetic (29th Yearbook)
The Teaching of Reading (Pt. I, 24th Yearbook)

(2) The Textbook

Language in Use for Third, Fourth, Fifth, and Sixth Grades and *English in Use* for Seventh Grade. Of this great series, one successful teacher told her board members who were considering ways and means of reducing costs of instruction, "If it were necessary, I could build all my courses, except arithmetic, around *Language in Use!*"

(3) The Pupil's Learning Power

The Instructional Tests in Arithmetic and in Reading by Sangren and Reidy and Wilson are to be used by the teacher throughout the year. We urge you to refrain from following the example of one superintendent who gave the tests all at one gulp with no attempt at instruction between tests. When used as an instructional tool they show quickly to the teacher the points at which pupils begin to fail so that the teacher can use her time to the best advantage.

(4) Educational Measurement Tests

For general pupil classification and diagnostic survey testing, use the widely standardized intelligence and achievement tests for grades and high school and college. Our catalogues describe many forms used for specific results. Write for them! *The Public School Achievement Tests* cover ten elementary school subjects and have four separate forms.

PUBLIC SCHOOL PUBLISHING COMPANY
509-11-13 North East Street
BLOOMINGTON, ILLINOIS